Psychometrika

CONTENTS

VOLUME TWO DECEMBER 1937 NUMBER FOUR

PSYCHOMETRIKA—VOL. 2, NO. 4
DECEMBER, 1937

A NOTE ON ROTATION OF AXES

CHARLES M. HARSH

Harvard University

Differences in magnitude among the communalities of a set of tests may influence the convergence of successive approximations to the limiting hyperplanes of the configuration. A slight alteration in technique is presented, which may be useful in avoiding this effect and perhaps hastening convergence.

In seeking simple structure in factor analyses the rotation of axes by a method of successive approximations is still a flexible and useful technique, although short-cut methods are preferable whenever they can be applied. The essential argument has been admirably presented by Thurstone*, but it is to be noted that in applying this method, (as, indeed, in applying most factorial methods), the exact technique used may depend upon non-mathematical considerations. With regard to the method outlined by Thurstone, the trial vector $\mathbf{\Lambda}$ is adjusted by means of a correction vector, determined by the projections of the non-factor tests on $\mathbf{\Lambda}$. Thus the tests are weighted by their communalities in determining the correction vector. There is no right or wrong about the logic, but the following alternative is presented as applicable when it is desired to give all tests equal weighting in determining the adjustment.

Referring to Figure 1, if there is only one test vector **J**, then the trial vector $\mathbf{\Lambda}$ can be adjusted to become orthogonal to **J** by means of a correction vector **S**, but contrary to the statement preceding equation 1), (ibid p. 65), this correction vector is *not* colinear with $-(\mathbf{\Lambda}\cdot\mathbf{J})\mathbf{J}$. **S** may be resolved into two components, **R** and **K**, of which **R** is colinear with $-\mathbf{J}$ and equal in length to v'_j, the projection of $\mathbf{\Lambda}$ on **J**.

Thus

$$\mathbf{R} = -\left(\frac{\mathbf{\Lambda}\cdot\mathbf{J}}{j}\right)\overline{\mathbf{J}} \equiv -v'_j\,\overline{\mathbf{J}}$$

where $\overline{\mathbf{J}}$ is a unit vector $\dfrac{\mathbf{J}}{j}$, and $j \equiv h_j$, the length of the test vector **J**.

*Thurstone, L. L., "The bounding hyperplanes of a configuration of traits," *Psychometrika*, 1936, **1**, **61-68**.

Letting $\overline{\mathbf{K}}$ be a unit vector orthogonal to $\overline{\mathbf{J}}$, and as the angle between **S** and **R** is $\frac{1}{2}(90 - \vartheta)$, it follows that

$$\mathbf{S} = -v'_j \overline{\mathbf{J}} + v'_j \tan \tfrac{1}{2}(90 - \vartheta) \overline{\mathbf{K}} .$$

Only when the angle of adjustment is very small can it be said that **R** is a fair approximation of **S**, so this method of adjustment seems inefficient. To approach the multi-test case, let **A′** be a vector sum, for all tests, of the correction components parallel to the test vectors, i.e., the sum of the projections of the trial vector Λ on the test vectors. Then, since **J** is not necessarily a unit vector,

$$\mathbf{A}' = \sum_j v'_j \overline{\mathbf{J}} = \sum_j \frac{v_j \mathbf{J}}{j^2} . \text{*}$$

If it is desired that **C** be as nearly as possible orthogonal to each

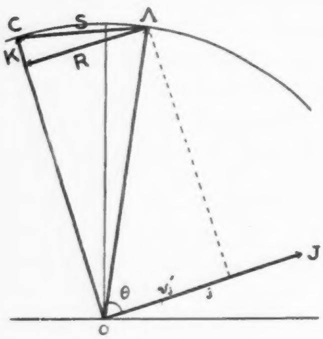


FIGURE I.

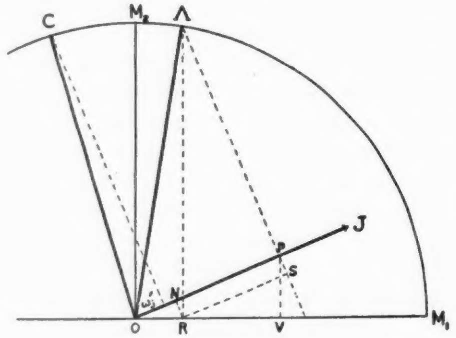


FIGURE II.

of the N_f non-factor tests, then **C** will be orthogonal to **A′**, and the **R** correction component will be $-\dfrac{\mathbf{A}'}{N_f}$. Such a correction gives equal consideration to the vector *direction* of each non-factor test, (i.e., to the factorial composition of the test, rather than to the magnitude of its communality). In contrast, Thurstone's **A** vector weights each test by its communality, first, by considering the projection of the test on the trial vector Λ, (instead of vice versa), and again, by multiplying this projection, v_j, by a direction vector **J**, rather than by a unit direction vector $\mathbf{J}/j = \overline{\mathbf{J}}$.

*Prime notations will be used to indicate symbols differing from those used by Thurstone, (ibid.). Thus $v'_j = v_{j/j}$.

Figure 2 represents a simple case in terms of two orthogonal reference axes, M_1 and M_2. To make **C** as nearly as possible orthogonal to any test **J**, we wish to minimize the projection of **C** on **J**, i.e.

$$w'_j = \frac{w_j}{j} = \frac{\mathbf{C}\cdot\mathbf{J}}{j} .$$

Following the reasoning of the above-mentioned article, $\mathbf{C} = m\mathbf{\Lambda} + p\mathbf{B}$, where **B** is a unit vector orthogonal to Λ in the plane of Λ and **J**. Then equation 22), (ibid), becomes

$$s^2\Sigma v'_j z'_j + s[\Sigma z'^2_j - \Sigma v'^2_j] - \Sigma v'_j z'_j = 0.$$

The only difference in solving the quadratic is the use of the v'_j and z'_j sums in place of the v_j and z_j sums. In terms of the reference axes the correction component along **J** is $v'_j\bar{\mathbf{J}} \equiv \overrightarrow{\mathbf{OP}}$. By constructing the various perpendiculars shown as dotted lines in Figure 2 it becomes apparent that

$$\overrightarrow{\mathbf{OP}} = \overrightarrow{\mathbf{ON}} + \overrightarrow{\mathbf{NP}} = (\lambda_{1p}\lambda_{j1})\bar{\mathbf{J}} + (\lambda_{2p}\lambda_{j2})\bar{\mathbf{J}},$$

or, in the general case involving more than two axes,

$$\overrightarrow{OP} = \sum_M (\lambda_{Mp}\lambda_{jM})\bar{\mathbf{J}} ,$$

where λ_{Mp} is the projection of the unit vector Λ on a reference axis **M**. But

$$\bar{\mathbf{J}} = \sum_M \lambda_{jM}\bar{\mathbf{M}} ,$$

consequently, $\overrightarrow{OP}$ may be considered in terms of its components $\overrightarrow{OV}$ and $\overrightarrow{VP}$, or in the general case, in terms of its projections on the various reference axes.

$$\therefore \quad \overrightarrow{OP} = \sum_M \left(\sum_M \lambda_{Mp}\lambda_{jM}\right)\lambda_{jM}\bar{\mathbf{M}} = \sum_M v'_j\lambda_{jM}\bar{\mathbf{M}} .$$

Summing for all the non-factor tests,

$$\mathbf{A}' = \sum_M \left[\sum_M v'_j\lambda_{jM}\bar{\mathbf{M}}\right] = \sum_M \left[\sum_j v'_j\lambda_{jM}\bar{\mathbf{M}}\right] .$$

This is the form in which **A′** is obtained on a work sheet, for each v'_j may be calculated in one continuous machine operation, and its products with the various λ_{jM} values tabulated in as many columns as there are reference axes M. The sum of each column gives an $a_{A'M}$, the component of **A′** along the corresponding reference axis. As the reference axes are orthogonal, it follows that

$$(\mathbf{A}')^2 = \sum_M \left(\sum_j v'_j\lambda_j^M\right)^2 = \sum_M (a_{A'M})^2 .$$

After substituting **A′** for **A** in Thurstone's equations 32) and 33), (ibid. p. 68), the rest of the procedure is the same as there outlined. Although its derivation may seem rather complex, this alternative method actually involves only one additional operation, namely dividing each v_j by the j value for the test.

This alternative method is especially applicable if one is more interested in the "primary factor" configuration than in the actual tests used. The low communality of a test may result from unreliability and may not warrant discounting the test when determining principal axes. An evaluation of **A′** depending upon the *direction* of the test vectors should give a somewhat more accurate estimation of the primary axes, and thus hasten the convergence of the successive approximations, **C**. The method has been used in various problems during the past year and seems to reduce the number of approximations necessary to yield a given degree of convergence by from one-fourth to one-half. This advantage should be most marked in cases involving a large number of tests with widely divergent communalities, whereas there should be no advantage if all tests had equal communalities.

PSYCHOMETRIKA—VOL. 2, NO. 4
DECEMBER, 1937

SAMPLING THEORY IN ITEM ANALYSIS

WALTER W. MERRILL, JR.

Since item values obtained by item analysis procedures are not always stable from one situation to another, it follows that selection of items for validity or difficulty is sometimes useless. An application of Chi Square to testing homogeneity of item values is made, in the case of the *UL* method, and illustrative data are presented. A method of applying sampling theory to Horst's maximizing function is outlined, as illustrative of author's observation that the results of item analysis by any of various methods may be similarly tested.

One finds little or no mention in psychometric literature of the application of sampling theory to item analysis. It is not necessarily true that an item analysis is justified wherever there are items, and it is the object of this paper to point out a method of testing the data to determine whether or not it justifies such an analysis and at the same time of putting an approximate valuation on the significance of the results.

Item analysis has as its object the selection of the best items of a group and the consequent elimination of the "deadwood". A fundamental, though seldom stated, assumption underlying an analysis is, therefore, that the items actually do differ from one another in some respect such as difficulty, or more commonly, validity. For purposes of discussion, we shall define items differing in either difficulty or validity as heterogeneous, those not so differing as homogeneous. We shall also speak of items as being heterogeneous with respect to only one of these characteristics, i.e., heterogeneous with respect to difficulty.

Unless one is reasonably sure that items are heterogeneous, he is not justified in going through an item analysis. It should furthermore be remembered that even though items are really homogeneous, sampling variation will cause them to differ from one another, especially if the number of cases to which each item is applied is small. The various methods of validating test items will tell one that a certain item is more valid than a certain other for the sample in question, but what one actually desires to know is whether or not it will continue to be more valid as it is given to more and more men. If the items are heterogeneous with respect to validity, one can say with some confidence that the most valid items in one sample will in general be the most valid in any other sample, and the use of the good items for predictive purposes is therefore justified. In the event, how-

ever, that there is a strong probability of the items being homogeneous, there is no justification for any selection.

One of the most commonly used methods of item analysis is the Upper-Lower Method(halves, thirds, or 27%), and this method lends itself particularly well to a test for heterogeneity with respect to both difficulty and validity. This test will prove with a reasonable degree of certainty that the items are heterogeneous but will never prove that they are homogeneous. It will merely indicate that they may be homogeneous, and it would seem that it is up to the investigator to show that they are heterogeneous before going through a complete analysis and then drawing positive conclusions from his results.

To test for heterogeneity with regard to difficulty $2 \times N$ contingency table is set up with two columns containing respectively total passes and total failures, and each row referring to one item. N is the number of items. A value for χ^2 is then obtained by the usual methods. Using $N—1$ as the number of degrees of freedom, one can find the probability of a value of χ^2 as large or larger than that obtained arising through mere chance by reference to a table such as Table III in R. A. Fisher's *Statistical Methods for Research Workers.** If P is less than .01, one is fairly safe in assuming that the data is heterogeneous, but if it is larger than say .10, it is certainly misleading to call it heterogeneous without strong *a priori* reasons for believing so. Between .01 and .10 it is up to the investigator to use his own judgment.

Since a valid item is one in which the upper group does well in reference to the lower group, a test for heterogeneity with regard to validity may be devised by setting up a similar $2 \times N$ contingency table with one column containing the sum of the numbers of upper group passes and lower group failures, and the other column containing the sum of the numbers of upper group failures and lower group passes. It will be noticed that when two items have the same validity as determined by the Upper-Lower Method, they give rise to identical entries in the $2 \times N$ contingency table. For instance, let us consider the items 13, 2, 11, 4 and 7, 8, 5, 10, where the four numbers represent respectively upper group passes, upper group failures, lower group passes and lower group failures. According to the Upper-Lower Method both of these items have validities of .13. Thus, if the test for heterogeneity of validities is applied, we find we have 17 and 13 for both items in the first and second columns of the contingency table.

*Fisher, R. A. *Statistical Methods for Research Workers.* London, Oliver and Boyd, 1934. Pages 110-111.

As an example of the use of this test, we shall consider data obtained in the testing of 45 mechanics in a large manufacturing organization. These men were first rated by their superiors and then given a number of performance tests, among which were the Minnesota Mechanical Assembly Tests (I and II). It was found that the correlation with the criterion could be materially increased by selecting only seven of the twenty items comprising this battery of items. After applying the χ^2 test described above, however, it was decided that, on the basis of the data at hand, there was no justification for the selection of any particular items.

The calculations involved are given in Tables I and II. In both tables, the first two columns, headed "Box" and "Item", are merely to identiy the items. In Table I, where we are testing for heterogeneity with respect to difficulty, the next column, headed "0", gives, for each item, the total number of both upper and lower group passing the item. The column headed "C" contains, for each item, the number we should expect to pass the item if there were no difference in difficulty between items. It is the mean of the figures in the "O" column. The column headed "D" is the difference between the "0" and the "C" columns, while the fourth column gives the value of χ^2 as calculated by the standard formula $\chi^2 = \frac{(0 - C)^2}{C}$. The next four columns are the same except that here we are dealing with the total number failing to pass each item. In the example given, we obtain a grand total χ^2 of 187.24. Since we are dealing with a 20×2 contingency table, we have 19 degrees of freedom, and using R. A. Fisher's table mentioned above with $n = 19$, we find P to be less than .01. This means that, were the data homogeneous with respect to difficulty, we should obtain differences between items as great as those observed in this sample, less than once in a hundred times. We can therefore conclude with some assurance that the items are heterogeneous with regard to difficulty.

Table II, in which we test for heterogeneity with respect to validity, is similar to Table I except that the two columns headed "0" are respectively the sum of the upper group passes and lower group failures and the sum of the upper group failures and lower group passes. In this case, we obtain a total χ^2 of only 8.61, and reference to the table indicates that differences in validity as great as those encountered would occur by pure chance in from 97 to 98 out of 100 such samples. We accordingly have no basis for declaring the data heterogeneous nor any justification for a selection of items.

The writer realizes that the application of this test to a problem involving a large number of items is as long a task as the item analysis itself. He believes, however, that its importance is such as to justify the extra time and expense. Should it happen that all the items are given to the same group of men, or even to the same number of men, the following alternative method of testing heterogeneity of validities is suggested.

This method involves a formula based on the validities of the various items and therefore cannot be applied until the item analysis is completed. Once the validities have been calculated, however, it becomes relatively simple to test them for heterogeneity.

The method consists of calculating χ^2 from the formula

$$\chi^2 = \frac{2\,N\,m\,\sigma^2}{1 - \bar{x}^2},$$

where N is the number of items, m is 27% (or half or one-third) of the number of subjects tested, σ is the standard deviation of the validities about their mean, and $\bar{x}$ is the mean validity. A χ^2 table is then used as before with $n = N - 1$. In using this method the investigator must go through the complete analysis and then perhaps find his work meaningless, but in the event that the validities do prove to be heterogeneous, the work is considerably less than that necessary to apply the test as given in the first part of the paper and then go ahead and calculate the validities.

The derivation of the formula $\chi^2 = \dfrac{2\,N\,m\,\sigma^2}{1 - \bar{x}^2}$ is as follows:

If we let $G = U_p + L_f$ and $B = U_f + L_p$ where U and L indicate upper and lower groups respectively and the subscripts p and f indicate passes and failures, we have

$$\chi^2 = \frac{\Sigma(G - \bar{G})^2}{\bar{G}} + \frac{\Sigma(B - \bar{B})^2}{\bar{B}},$$

provided that the upper and lower groups are both the same in number for every item (i.e., do not vary in number from item to item). If $m = 27\%$ (or half or one-third) of the number of subjects to which each item is applied, we have

$$U_p + U_f = m = L_p + L_f,$$

whence

$$G = m + (U_p - L_p),$$

and

$$B = m - (U_p - L_p).$$

Furthermore if x be the validity of an item,

$$x = \frac{U_p - L_p}{m}. \quad \text{arbitrary}$$

Thus

$$G = m + mx,$$

and

$$B = m - mx,$$

and

$$\chi^2 = \frac{\Sigma\left(m + mx - \frac{\Sigma(m+mx)}{N}\right)^2}{\frac{\Sigma(m+mx)}{N}} + \frac{\Sigma\left(m - mx - \frac{\Sigma(m-mx)}{N}\right)^2}{\frac{\Sigma(m-mx)}{N}}$$

$$= \frac{\Sigma\left(m + mx - m - \cancel{m}\frac{\Sigma \cancel{m}x}{N}\right)^2}{m\left(1 + \frac{\Sigma x}{N}\right)} + \frac{\Sigma\left(m - mx - m + m\frac{\Sigma x}{N}\right)^2}{m\left(1 - \frac{\Sigma x}{N}\right)}$$

$$= \frac{m^2\,\Sigma(x - \bar{x})^2}{m(1 + \bar{x})} + \frac{m^2\,\Sigma(\bar{x} - x)^2}{m(1 - \bar{x})}$$

$$= m\,N\,\sigma^2\left(\frac{1}{1 + \bar{x}} + \frac{1}{1 - \bar{x}}\right)$$

$$= \frac{2\,m\,N\,\sigma^2}{1 - \bar{x}^2}.$$

It should be noted that whereas the techniques given here are applicable only to the Upper-Lower Method, the underlying principles are applicable to any of the methods of item analysis. It appears important that those making frequent use of other methods should develop similar techniques for testing for heterogeneity of items. As an example of a method quite dissimilar to the Upper-Lower Method, we might take that described in Dr. Paul Horst's paper, "Item Selection by Means of a Maximizing Function".*

*Horst, Paul, "Item Selection by Means of a Maximizing Function", *Psychometrika*, 1936, **1**, pp. 229-244.

In this method, we consider two functions

$$u_k = \Sigma Y_k - M_y N_k$$

$$v_k = \Sigma S_k - M_s N_k$$

where M_y = mean of criterion scores,
M_s = mean of total test scores,
ΣY_k = the sum of the criterion scores for all who answered item k correctly.
ΣS_k = the sum of the total test scores for all who answered item k correctly.
N_k = the number who answered item k correctly.

If m items are to be selected, the selection process then consists of choosing the m which maximize the function

$$\frac{\Sigma u_k}{\sqrt{\Sigma v_k}}.$$

Without going into the details of the solution of this problem, it can be easily seen that it consists of selecting items having certain values of u and v and eliminating other items having different values of u and v. It at once becomes obvious then that the use of this method is based upon the assumption that the u and v values differ from item to item by something more than chance sampling variation.

As a first step in developing a test to indicate whether these pairs of u and v values differ significantly from one another, let us examine the functions u and v more closely. Let us write them in the form

$$u_k = N_k \left(\frac{\Sigma Y_k}{N_k} - M_y\right),$$

$$v_k = N_k \left(\frac{\Sigma S_k}{N_k} - M_s\right),$$

and to simplify our discussion let us set

$$\left(\frac{\Sigma Y_k}{N_k} - M_y\right) = \overline{Y}_k,$$

and

$$\left(\frac{\Sigma S_k}{N_k} - M_s\right) = \overline{S}_k.$$

It is now at once apparent that differences in the values of u and v from item to item may be due to differences in either N_k or $\overline{Y}_k$ and $\overline{S}_k$.

N_k is by definition simply a measure of the difficulty of item k, so that any variation in validity affecting u and v will show up as a variation in $\overline{Y}_k$ and/or $\overline{S}_k$ from item to item. If we are chiefly interested in choosing valid items, a test for heterogeneity of the values of u and v is of little help, since this heterogeneity can well be caused by differing difficulties rather than differing validities.*

It is a simple matter, however, to consider N_k, $\overline{Y}_k$, and $\overline{S}_k$ rather than u_k and v_k in testing for heterogeneity. If the values of N_k differ significantly, we conclude that there is a significant difference in difficulty from item to item, and if the values of $\overline{Y}_k$ and $\overline{S}_k$ differ significantly, we can conclude that there is a significant difference in validity from item to item.

Dr. S. S. Wilks has furnished the author with a method for testing these quantities for heterogeneity. This test proceeds as follows:

If N_k, $\overline{Y}_k$, or $\overline{S}_k$, is a function of k, i. e., their variations from item to item are caused by something more than chance sampling error, we should get a significant correlation between respective values for each item when the test or total group of items is given to two random samples of subjects drawn from the same universe or population.

To put this another way, suppose we give the items of a test to a group of subjects G_1 and then plot the values of N_k along a line in order of magnitude. Then suppose we give the same test to another group G_2, drawn from the same universe and again plot the values of N_k. If nothing but chance is operating, our second set of points should be a random redistribution of the first set. If, however, there is really a significant difference in difficulty from item to item, our second set of points should line up on our line in substantially the same order as the first set. To test for heterogeneity of difficulties, then, our procedure would consist of breaking up our original group of subjects into two random halves and then calculating the correlation between the values of N_k for the two groups. If the correlation differs significantly from zero,† we can assume heterogeneity as regards difficulty.

The test for validity is similar. Here, however, we are considering two variables, $\overline{Y}_k$ and $\overline{S}_k$. If we give our test to two groups of sub-

*While this fact is very important in devising a test for significance, it is of little consequence in reference to the actual selection process as developed by Dr. Horst. Roughly speaking, he chooses large values of u and small values of v so that the effect of difficulty is nullified.

†For a method of testing the significance of a correlation coefficient, see R. A. Fisher, (*op. cit.*), page 179.

jects, the values of $\overline{Y}_k$ and $\overline{S}_k$ for the two groups should distribute themselves in the same manner in the $\overline{Y}\overline{S}$ plane. If the values for the second group are simply a random redistribution of the points for the first group, we cannot assume anything other than chance to be operating. While we could calculate correlations between values of $\overline{Y}_k$ and between values of $\overline{S}_k$ for the two groups as in the case of N_k, a more exact test in this case is as follows:

Suppose we consider $\overline{Y}_k'$, $\overline{S}_k'$, $\overline{Y}_k''$, and $\overline{S}_k''$, where the single primes refer to the first group and the double primes to the second group. Let

$$U = \frac{R}{\begin{vmatrix} 1 & r_{\overline{y}_k' \overline{y}_k''} \\ r_{\overline{y}_k' \overline{y}_k''} & 1 \end{vmatrix} \cdot \begin{vmatrix} 1 & r_{\overline{s}_k'' \overline{s}_k'} \\ r_{\overline{s}_k' \overline{s}_k''} & 1 \end{vmatrix}},$$

where R is the fourth order determinant of the intercorrelations of the four variables. If only chance variation is present, the value of U should approximate unity, and the smaller the value of U, the greater the confidence with which we can say that mere chance is not causing the variation. Under the assumption of chance variation and normality of distribution of the four variables, the probability that U will be less than a given value, say U_0, is

$$U_0^{\frac{N-4}{2}} (N - 3 - (N-4) U_0),$$

where N is the total number of items. If the probability as calculated by this formula is less than say .01, we are reasonably safe in rejecting the chance hypothesis and concluding that the selection of items by the u, v method is really valid for the test in question.

TABLE I
Test for heterogeneity of item difficulties.

Box	Item	Upper + Lower Pass 0	C	D	χ^2	Upper + Lower Fail 0	C	D	χ^2
I	1	28	23	5	1.09	2	7	5	3.57
	2	29	23	6	1.56	1	7	6	5.14
	3	30	23	7	2.13	0	7	7	7.00
	4	29	23	6	1.56	1	7	6	5.14
	5	12	23	11	5.26	18	7	11	17.28
	6	11	23	12	6.26	19	7	12	20.28
	7	14	23	9	3.52	16	7	9	11.57
	8	26	23	3	.39	4	7	3	1.28
	9	6	23	17	12.55	24	7	17	41.28
	10	29	23	6	1.56	1	7	6	5.14
II	1	29	23	6	1.56	1	7	6	5.14
	2	27	23	4	.69	3	7	4	2.28
	3	23	23	0	0	7	7	0	0
	4	25	23	2	.17	5	7	2	.57
	5	16	23	7	2.13	14	7	7	7.00
	6	18	23	5	1.09	12	7	5	3.57
	7	24	23	1	.04	6	7	1	.14
	8	27	23	4	.69	3	7	4	2.28
	9	28	23	5	1.09	2	7	5	3.57
	10	26	23	3	.39	4	7	3	1.28
Total		457	460		43.73	143	140		143.51

$\chi^2 = 43.73 + 143.51 = 187.24.$

TABLE II
Test for heterogeneity of item validities.

Box	Item	Upper pass + lower fail 0	C	D	χ^2	Upper fail + lower pass 0	C	D	χ^2
I	1	13	16	3	.56	17	14	3	.64
	2	16	16	0	0	14	14	0	0
	3	15	16	1	.06	15	14	1	.07
	4	14	16	2	.25	16	14	2	.29
	5	17	16	1	.06	13	14	1	.07
	6	18	16	2	.25	12	14	2	.29
	7	15	16	1	.06	15	14	1	.07
	8	13	16	3	.56	17	14	3	.64
	9	17	16	1	.06	13	14	1	.07
	10	14	16	2	.25	16	14	2	.29
II	1	16	16	0	0	14	14	0	0
	2	18	16	2	.25	12	14	2	.29
	3	16	16	0	0	14	14	0	0
	4	18	16	2	.25	12	14	2	.29
	5	19	16	3	.56	11	14	3	.64
	6	19	16	3	.56	11	14	3	.64
	7	15	16	1	.06	15	14	1	.07
	8	16	16	0	0	14	14	0	0
	9	15	16	1	.06	15	14	1	.07
	10	15	16	1	.06	15	14	1	.07
Total		319	320		3.91	281	280		4.70

$\chi^2 = 3.91 + 4.70 = 8.61.$

PSYCHOMETRIKA—VOL. 2, NO. 4
DECEMBER, 1937

A METHOD OF FACTOR ANALYSIS BY MEANS OF WHICH ALL COORDINATES OF THE FACTOR MATRIX ARE GIVEN SIMULTANEOUSLY

PAUL HORST
The Procter and Gamble Company, Cincinnati, Ohio

In general, the methods of factor analysis developed during the past five years are based on the reduction of the correlational matrix by successive steps. The first factor loadings are determined and eliminated from the correlational matrix, giving a residual matrix. This process is continued for successive factor loadings until the elements of the last obtained residual matrix may be regarded as due to chance. The method outlined in this paper assumes the maximum number of factors m in the correlational matrix. The m factor vectors are solved for simultaneously. Once the m factor vectors are found, any vectors having only negligible factor loadings may be discarded.

One of the greatest obstacles to the application of factor analysis techniques in psychology by those who are familiar with the theory underlying the techniques is the great amount of arithmetical work involved. Even the most rapid methods of multiple factor analysis thus far presented become laborious when the number of variables exceeds twenty or thirty.

Most of these methods take as the basic data the matrix of zero order correlations between the variables. Assuming that this matrix has n variables the problem is to find that matrix of n rows and $(m<n)$ columns which, when post-multiplied by its transpose, most nearly approximates the correlational matrix. That is, if we let

$r_0 =$ a correlational matrix of n variables,

$a =$ a matrix of n rows and m columns,

the problem is to determine a so that the approximation

$$a\,a' \doteq r_0 \tag{1}$$

will be satisfactorily close.

The best known methods for determining a begin by first determining a single column matrix or vector which, according to some criterion, gives a satisfactory approximation in (1). Designating this

vector by a_1 the usual procedure is to determine the residual matrix r_1 in the equation

$$r_1 = r_0 - a_1 a_1'. \tag{2}$$

Similarly a second column vector a_2 is determined, such that the approximation

$$a_2 a_2' \doteq r_1 \tag{3}$$

is satisfactorily close. This procedure is carried on until the magnitudes of the elements in the last residual matrix are considered so small as not to differ significantly from zero. The matrix a in (1) is then taken as the matrix whose columns are these successive column vectors.

The methods involving operations of this type vary considerably in the amount of arithmetical work required and in the degree of approximation achieved. However, even the most rapid methods are prohibitive in the case of say 100 variables.

Doubtless the scope and usefulness of multiple factor methods would be greatly increased if feasible arithmetical methods were available for handling large numbers of variables. Several obvious applications might be mentioned. In the field of personality, interest, and attitude measurement it is not uncommon to find more than 100 items on schedules or inventory forms purporting to evaluate the non-intellective aspects of human behavior.

Again let us consider the subject of item selection techniques. It is known that the degree of refinement to which it is profitable to carry these techniques is definitely limited by the stability of the responses to the individual items composing a test battery. Let us assume that we have a matrix in which the rows represent persons and the columns items. Each element in this matrix represents a given person's response to a given item. We assume that a response is either correct or incorrect, so that all elements in the matrix are either unity or zero. A matrix of this type, explicit or implied, together with a set of criterion scores constitutes the basic data from which item selection techniques proceed. However, most matrices of this type, representing experimental data, are more or less unstable and in general the item selection techniques capitalize on the errors incident to this instability, thus building up a spurious validity for the test composed of the selected items.

Suppose now that the experimental test battery consists of 500 items and that we have at least this many cases. In general the rank

of the matrix would be 500. It would be discouraging and thus far, at least, unnecessary to assume that the 500 items represent 500 different abilities. Judging from studies already carried out, it is not unreasonable to assume that the total number of distinct significant human abilities and traits does not exceed 20 or 30.

Granting this assumption, it should be possible to approximate the matrix of rank 500 with one of rank 30 whose elements would be enormously more reliable than those in the former matrix. It should not be difficult to develop an item selection technique to operate on this matrix of rank 30 so that the spuriousness in the validity of the test consisting of the selected items would be negligible.

Obviously, of course, this suggested procedure involves a factor-analysis of 500 variables and would not be feasible with current methods.

Other applications for more rapid factor-analysis techniques might be mentioned but these will suffice to suggest their desirability. We shall therefore outline a procedure which is straight-forward and comparatively rapid.

To make our treatment as general as possible we begin by considering a matrix z whose columns represent variables and whose rows represent individual cases in a population. In particular the variables may be test scores and the cases may be people. We specify the number of columns in z by n and the number of rows by N. Without loss of generality we may assume $n < N$.

Eckart and Young (**1**) have pointed out in somewhat different notation that if z is a real matrix, it may always be expressed in the form

$$z = H\, k^{\frac{1}{2}}\, h', \tag{4}$$

where

H is a matrix having the same number of rows and columns respectively as z and satisfying the equation

$$H'\, H = I. \tag{5}$$

$k^{\frac{1}{2}}$ is a diagonal matrix of order n.
h' is an orthogonal matrix of order n.

From (4) and (5) we get

$$z'\, z = h\, k\, h', \tag{6}$$

or

$$z'\, z\, h = h\, k. \tag{7}$$

In equation (7) it is known that the diagonal elements of k are the roots of $z' z$. Without loss of generality we may assume that from upper left to lower right the roots are arranged in descending order of magnitude. If we substitute zeros for all but the first m diagonal elements of k and designate the resulting matrix by k_m, Eckart and Young have shown that the matrix of rank m which best approximates z in the least square sense is given by

$$z \doteq H\, k_m^{\frac{1}{2}}\, h'. \tag{8}$$

Furthermore the matrix of rank m which best approximates $z' z$ in the least square sense is

$$z' z \doteq h\, k_m\, h'. \tag{9}$$

Suppose now we let

$$a = h_m\, k_m^{\frac{1}{2}}, \tag{10}$$

where h_m means that we have dropped all but the first m columns of h. From (9) and (10) we have

$$a\, a' \doteq z' z, \tag{11}$$

and from (10)

$$a' a = k_m. \tag{12}$$

The matrix a has been designated the "principal axis" or "principal component" resolution of $z' z$. The mathematical solution for the column vectors of a is well known. If we let k_i be the *i'th* root of $z' z$ and a_i the *i'th* column vector of a, the following relation holds:

$$(z' z - k_i\, I)\, a_i = 0. \tag{13}$$

From equation (13), $(n-1)$ of the coordinates of a_i can be solved for as ratios of the remaining coordinate. The actual values of these coordinates are then determined from the relation

$$a_i' a_i = k_i. \tag{14}$$

The mathematical solution requires, however, that the root of $z' z$ be first found. The solution for the roots is laborious for more than 5 or 6 variables and solutions of the type indicated in (13) are tedious with more than 15 or 20 variables. For this reason the methods previously mentioned, involving the progressive reduction of $z' z$ have been developed.

Suppose, however, that we have some reasonable assurance that $z' z$ can be adequately approximated by a matrix of rank no greater

than m where m is known. We consider any arbitrary matrix a_1 of n rows and m columns, and an unknown matrix a_2 also of n rows and m columns. Then in the approximation equation

$$a_1\, a_2' \doteq z'\, z \tag{15}$$

our problem is to determine a_2 so that (15) will give the best approximation in the least square sense. It can be shown that the least square solution for a_2 proceeds as follows: Premultiplying both sides of (15) by a_1' we have

$$(a_1'\, a_1)\, a_2' \doteq a_1'\, z'\, z. \tag{16}$$

Premultiplying again by the inverse of $a_1'\, a_1$ we have

$$a_2' = (a_1'\, a_1)^{-1}\, a_1'\, z'\, z. \tag{17}$$

Next we take an unknown matrix a_3 in the equation

$$a_2\, a_3' \doteq z'\, z. \tag{18}$$

Analagous to (17) the least square solution for a_3' is

$$a_3' = (a_2'\, a_2)^{-1}\, a_2'\, z'\, z. \tag{19}$$

Then (18) will be at least as good an approximation as (15) and in all but certain theoretical cases it will be better. In the general case we have

$$a'_{i+1} = (a_i'\, a_i)^{-1}\, a_i'\, z'\, z. \tag{20}$$

We continue this procedure, each time getting a closer approximation to $z'\, z$. When we have the best approximation obtainable within the limits of the number of decimals carried in the solution we shall have

$$a_{i+1} = a_{i-1}. \tag{21}$$

Therefore

$$a_{i-1}\, a_i' \doteq z'\, z \tag{22}$$

is the best approximation obtainable in the least square sense. But equation (11), where a is defined by (10), is also the best approximation in the least square sense. Therefore if we let

$$a_{i-1} = b_1,\ a_i = b_2, \tag{23}$$

we have from (11) and (22)

$$a\, a' = b_1\, b_2'. \tag{24}$$

Our problem now is to solve for a in (24).

First we consider two matrices c and e of order m such that

$$a = b_1 \, c \, u, \tag{25}$$

$$a = b_2 \, e \, u, \tag{26}$$

where u is an orthogonal matrix as yet unspecified.

From (24), (25) and (26) we have

$$a \, a' = b_1 \, c \, u \, u' \, e' \, b_2',$$

or since u is orthogonal

$$a \, a' = b_1 \, c \, e' \, b_2'. \tag{27}$$

From (24) and (27)

$$b_1 \, b_2' = b_1 \, c \, e' \, b_2'. \tag{28}$$

Hence

$$c \, e' = I,$$

or

$$e = c'^{-1}. \tag{29}$$

Therefore, we may write equation (26)

$$a = b_2 \, c'^{-1} \, u. \tag{30}$$

From (25) and (30) we get

$$b_1 \, c = b_2 \, c'^{-1}, \tag{31}$$

or

$$b_1 \, c \, c' = b_2. \tag{32}$$

Premultiplying both sides of (32) by b_1' we have

$$b_1' \, b_1 \, c \, c' = b_1' \, b_2, \tag{33}$$

and premultiplying (33) by $(b_1' \, b_1)^{-1}$ gives

$$c \, c' = (b_1' \, b_1)^{-1} \, b_1' \, b_2. \tag{34}$$

Equation (34) does not give a unique solution for c. In fact there is an infinite number of solutions for c. However, all of these solutions differ from one another only by an orthogonal transformation. Let us substitute in equation (30) any solution c which satisfies (34). Multiplying each side of (30) by its transpose we have

$$a' \, a = u' (c^{-1} \, b_2' \, b_2 \, c'^{-1}) u. \tag{35}$$

Substituting from (12) in the left-hand side of (35) and premultiplying by u we have

$$u\,k_m = (c^{-1}\,b_2'\,b_2\,c'^{-1})\,u. \tag{36}$$

Since k_m is a diagonal matrix and the quantity in parentheses is known, (36) enables us to solve for u. The solution for a in (30) then is given by the steps just outlined. In many problems, however, we are not interested in the solution for a as such but rather in some orthogonal transformation of a such that the coordinates of the transformed matrix will satisfy certain criteria. Therefore, in general we may write (30) in the form

$$a\,u' = b_2\,c'^{-1}, \tag{37}$$

and make our transformation on au' rather than a, thereby avoiding the necessity of solving for and multiplying by u'.

Our first practical consideration is the rapidity with which the solutions represented by equation (20) stabilize. This will depend entirely on the choice of the matrix a_1 in (15). If we could choose a_1 so that

$$a_1 = a\,g, \tag{38}$$

where g is any real non-singular matrix of order m, the approximation process would stabilize with a_2. This can be shown by substituting ag for a_1 in (17).

In general, of course, it is not possible to choose a_1 in this manner. But the closer we can come to so choosing a_1, the more rapidly will the successive approximations stabilize. Therefore, let us consider the problem of choosing a_1 to satisfy as nearly as possible equation (38). Let us assume that we have taken m large enough so that the approximation in (11) is reasonably close, that is, so that the residuals in the matrix ε given by

$$\varepsilon = z'\,z - a\,a' \tag{39}$$

are in general negligible as compared with the elements $z'\,z$ or aa'. The first m columns of aa' may be indicated by aa_m' where a_m is a matrix of order m made up of the first m rows of a. If we indicate the first m rows of $z'\,z$ by r_m we have

$$a\,a_m' \doteq r_m, \tag{40}$$

where the approximation is not necessarily in the least square sense but should in general approximate it closely. The left hand side of

(40) is precisely a_1, in (38) if we let g equal a_m'. Therefore r_m, or the first m columns of $z' z$ may be taken as a_1. And if these m columns are so selected that the intercorrelations of the variables represented by them are low, the successive approximations should stabilize more rapidly than if these intercorrelations are high.

Furthermore, we may select m sets of columns from $z' z$ adding corresponding row elements within a set to get each of the m columns of a_1. This procedure may be justified as follows: Without loss of generality we assume that the m sets of columns from $z' z$ are selected in order from left to right. The first column vector of a_1 will be the sum of the first s column vectors of $z' z$, the next column vector of a_1 will be the sum of the next s column vectors of $z' z$ and so on. Assume now that the first column vector of g in (38) is the sum of the first s row vectors of a and so on. Then a_1 made up from successive sets of s columns each taken from $z' z$ would approximate to ag where g is made up as indicated.

To prove this we consider again equation (11).

$$a\, a' \doteq z' z. \tag{11}$$

We let P_{s_1} be a matrix consisting of the first s column vectors of $z' z$, P_{s_2} a matrix consisting of the next s column vectors of $z' z$ and so on. If now we let a_{s_1} be the matrix of the first s row vectors of a, a_{s_2} the matrix of the next s row vectors of a and so on, then the first s column vectors of aa' will be given by aa'_{s_1}, the next s column vectors of aa' by aa'_{s_2} and so forth.

Therefore we have

$$a\, a'_{s_1} \doteq P_{s_1},$$

$$a\, a'_{s_2} \doteq P_{s_2},$$

or in general

$$a\, a'_{s_i} \doteq P_{s_i}. \tag{41}$$

We let the symbol (1) indicate a column vector all of whose coordinates are unity. If we post-multiply both sides of (41) by this vector we have

$$a\,(a'_{s_i} 1) \doteq P_{s_i} 1, \tag{42}$$

where the vector (1) is of order s. Obviously the right side of (42) is simply a vector whose coordinates are the sums of corresponding row coordinates of the matrix P_{s_i}.

On the left hand side the term in parentheses is a vector which is the sum of the *i'th* set of s row vectors of a. Thus we justify the procedure of summing successive sets of column vectors to get a_1.

The foregoing discussion shows, however, that the sets need not necessarily be successive or even have the same number of columns. As a matter of fact, when this method is used the most desirable arrangement is that all variables corresponding to columns within a set have as high intercorrelations as possible, and all variables represented by columns not in the same set have as low intercorrelations as possible.

The combination of columns in this manner is particularly well adapted to problems involving the factor analysis of individual items in a test battery. Here the variables to be combined to form the individual column vectors of a_1, will be all of the items in a specific type of subtest. For example, all vocabulary items may be combined, all arithmetical items, etc. In general the method is adapted to all large groups of variables which can be divided into logical sub-groups.

Assuming now that the successive approximations have been carried out until they are considered sufficiently stable, our next problem is to determine c in equation (34). As we have indicated previously, an infinite number of matrices exist which satisfy (34). Any one of these matrices are adequate for our purpose. Therefore, we shall take a solution for c which is readily obtainable. If in (34) we let

$$(b_1' \, b_1)^{-1} \, b_1' \, b_2 = S, \tag{43}$$

equation (34) may be written

$$c \, c' = S. \tag{44}$$

The matrix S is symmetrical. In fact any matrix of the form $(a_i' \, a_i)^{-1} \, a_i' \, a_{i+1}$ is symmetrical as can be shown by substituting the right hand side of equation (20) for a_{i+1}. We take as our clue for the solution of c a method of factor analysis discussed by Thurstone (4). This method is designated by Thurstone as the diagonal method. The factor matrix resulting from this method has all zero elements above the diagonal. Thurstone discusses the method only in connection with a correlational matrix. It is equally applicable, however, for factoring any matrix which is the product of any real matrix by its transpose. The matrix S in (42) is clearly of this type.

In considering the diagonal solution for c we shall first discuss the solution of a set of normal equations by the Doolittle method

(3). We regard S as the matrix of a set of normal equations. The Doolittle method consists of obtaining a set of reduced equations from the normal equations. The matrix of the reduced equations is of the diagonal form. That is, the first equation is a linear function of all the unknowns, the second, of all but the first unknown, and so on to the last, in which only the last unknown appears. These equations are the basis for what is called the back solution for the unknowns. Therefore, in the Doolittle method all elements below the diagonal in the matrix of the back-solution equations are zero. Furthermore, all diagonal elements are (-1). In the conventional Doolittle procedure, however, a preliminary set of reduced equations is derived in the process of obtaining the back-solution equations. Each row in the matrix of the preliminary set is proportional to the corresponding row of the back-solution matrix.

If we let

$v =$ the preliminary reduced matrix

$D =$ a diagonal matrix whose diagonal elements are also the diagonal elements of v.

$w =$ the back-solution matrix

we have

$$-D^{-1} v = w. \tag{45}$$

In conventional multiple regression problems where the Doolittle method is employed a column of constants must be carried along in the solution. However, we may carry out a Doolittle solution on the matrix S without the column of constants. In this solution it can be proved that

$$(v' D^{-\frac{1}{2}}) (D^{-\frac{1}{2}} v) = S. \tag{46}$$

From (44) and (46) we have then

$$c\, c' = (v' D^{-\frac{1}{2}}) (D^{-\frac{1}{2}} v), \tag{47}$$

and from (47)

$$c = v' D^{-\frac{1}{2}}, \tag{48}$$

since c may be any matrix which, when post multiplied by its transpose gives S.

From equations (25) and (30) we have

$$a\, u' = b_1\, c, \tag{49}$$

$$a\, u' = b_2\, c'^{-1}. \tag{50}$$

Thus knowing b_1 and c we can solve for au' by means of (47). If the successive approximations have been carried out to the point of stabilization within the limit of the number of decimals carried, (49) and (50) will give the same solution for au' therefore it will not be necessary to use (50) and consequently c^{-1} need not be determined. But in many cases one may be content to stop with an approximation a_{k+1} let us say, before complete stability is achieved. Then we should have

$$S = (a_k' a_k)^{-1} a_k' a_{k+1}, \tag{51}$$

and the approximate equations

$$a u' \doteq a_k c, \tag{52}$$

$$a u' \doteq a_{k+1} c'^{-1}. \tag{53}$$

In general (53) would give a better approximation to au' than (52), therefore it would be necessary to evaluate c^{-1}. This can be done most conveniently as follows: From equations (45) and (48)

$$c = - w' D^{\frac{1}{2}}, \tag{54}$$

and

$$c^{-1} = - D^{-\frac{1}{2}} w'^{-1}. \tag{55}$$

The inverse of c can be more conveniently calculated from (55) than from (48). This is true because the inverse of w is more readily calculated than the inverse of v, since all diagonal elements in w are (—1).

The method of factor analysis which we have outlined was employed on a correlational matrix of 17 variables. The centroid method of Thurstone and the iteration method of Hotelling (2) were also employed. It was found that the product of $a_2\, a_3'$ approximated the correlational matrix as closely as the centroid method and required only about three-fourths as much time for calculation. To get as close an approximation as was given by the iteration method of Hotelling it was necessary to carry the approximation to a_8. However, even this required only about one-fifth the time of the Hotelling method. The statistical assistants who made the computations were equally familiar with all three methods.

The detailed steps involved in carrying out the calculations for the new method will be discussed in a subsequent paper.

REFERENCES

1. ECKART, C, and YOUNG, G., "The Approximation of One Matrix by Another of Lower Rank", *Psychometrika,* 1936, **1**, pp. 211-218.
2. HOTELLING, H., "Analysis of a Complex of Statistical Variables into Principal Components", *Journal of Educational Psychology,* 1933, pp. 432-434.
3. MILLS, F. C., *Statistical Methods,* Henry Holt and Company, 1924, pp. 577-581.
4. THURSTONE, L. L., *The Vectors of Mind,* University of Chicago Press, 1935, pp. 78-81.

PSYCHOMETRIKA—VOL. 2, NO. 4
DECEMBER, 1937

THE SCALING OF PRACTICE DATA*

HERBERT WOODROW
University of Illinois

It is shown that in certain cases practice data approximately meet the assumptions involved in Thurstone's method of absolute scaling. An application of the method was accordingly made in the case of four test performances practiced for 39 days by a group of 56 subjects. The manner in which the practice data were scaled is described by using the data on practice in anagrams as an illustration. Scaling had little effect upon the correlations between initial and final score, but produced marked changes in the apparent effect of practice upon individual differences and in the correlations between initial score and gain.

A method of scaling which naturally suggests itself for use with practice data is that employed by Thurstone for studying the growth curve in ability. It is not apparent why a method which can be employed for that purpose might not also be applicable to practice data. It has even been alleged **(1)** that "A mental growth curve is at best a practice curve obtained in the absence of controlled conditions".

Thurstone's method, in its present form, is based on the assumption of a normal distribution of ability. As yet there appears to be no way of rigorously proving the validity of this assumption. A normal distribution of raw scores of course would not establish, nor would a skewed distribution thereof disprove, the assumption of a normal distribution of ability, since the relation between score and ability is not necessarily a linear one. However, in the case of two overlapping distributions on a scale measuring throughout the same trait, if the distribution of ability is normal for both the distributions in question, a certain mathematical criterion, established by Thurstone **(3)**, must be satisfied; and if this criterion is not satisfied, one has proof that both the distributions cannot be normal. The criterion consists simply in a linear relationship between the sigma values of the overlapping distributions, these sigma values (x-values) being those of the percentages of scores better than (or worse than) specified scores, when the x-values are calculated on the assumption of a normal distribution.

*This is one of a series of studies of the application of scaling methods to problems of experimental psychology in which the data are obtained from but a single group of subjects. For other studies, see **(4)**, **(5)**, **(6)**, and **(8)**.

Since the data at hand were obtained from practice experiments with two groups numbering only 56 and 82, respectively, it seemed desirable to increase the reliability of the proportion of scores falling below any specified value by averaging the proportions for several successive days, after first eliminating the effect of practice during those several days. In order to eliminate the effect of practice, the scores for each of the days in one short segment of practice were multiplied throughout by whatever number was necessary to make the mean score for each of them equal to the mean for all of them, that is, by the mean of the means for all the days included in the segment divided by the mean for the day in question. This correction not only resulted in equating the means for the several successive days but also gave daily σ's that differed very little from each other, though the variation may possibly in some cases have been slightly greater than could be accounted for satisfactorily purely on the basis of sampling of responses. Thus, in the case of horizontal adding, with 56 subjects, the σ's of the corrected scores for the first five days (which showed an average intercorrelation of .75) were 13.4, 11.3, 13.4, 12.6, and 12.1 None of these values differ seriously from the σ of the pool which was 12.6, nor from their arithmetical mean, namely, 12.5. And for the three final days the σ's of the corrected scores were 22.2, 22.4, and 22.5, while the σ of the pool was 22.4, as was also the mean of the three daily σ's. In view of these results, proportions calculated from a pool of several days may be regarded simply as representing more reliably than would those for any one day the proportions existing at that stage of practice at which the group mean equalled the mean of the pooled days. Since the reliability of the proportions depends in part upon the reliability of the scores, and the reliability of the scores is lower at the initial than at the final stages of practice, the number of days pooled in the initial segment of practice was larger than that pooled in the final segment. The number of days chosen for a single pool was varied so that the value $nr/(1 + (n - 1)r)$ was near +.94, r being the average of the test intercorrelations for the pooled days. This procedure rendered certain that the reliability of mean initial scaled scores would turn out to be high and approximately equal to that of final mean scaled scores, a result eminently desirable if the scores are to be used for the calculation of gains or the correlation between gains and initial scores. The pool representing initial ability, accordingly, included all the scores made on the first five days (except in the case of the modified spot-pattern test, in which case it included only the first three days) ; the pool represent-

ing final ability included the scores made on the last three days (days 37-39 in the only cases in which scaling was found feasible); and a pool representing intermediate ability also included three days. The proportions calculated from these pools are regarded as giving a more reliable picture, than those calculated from a single day, of the true distribution, not of true scores, but of scores having that degree of fallibility which actually characterized the scores at the stage of practice in question. It will be borne in mind, however, that true scores would show a smaller σ than these fallible scores. The decrease in the σ of the true scores as compared with that of the fallible scaled scores will be greater in the case of initial than in the case of final scores. It follows that the relation between the σ's of final and initial ability measured by true scores (likewise, by equally reliable scores) will differ from that between the σ's of the final and initial pool.

The three overlapping distributions, representing respectively initial, intermediate, and final stages of practice, were constructed for each of five tests practiced by 56 subjects for 39 sittings, and for each of four tests practiced by 82 subjects for 66 sittings. In the case of each test, two plots showing the relationship of the x-values were examined; one, the plot for the initial and intermediate distributions; the other, that for the intermediate and final distributions. In none of the tests practiced by the group of 82 subjects did both plots approximate a straight line. The explanation lies in part at least in the fact that the score distributions for this group in all cases showed a slight bimodality. In the case of the group of 56 subjects, however, both plots were very nearly linear in four of the five tests which were studied. The pooled distributions for these four tests were therefore scaled by Thurstone's method of absolute scaling. The tests were the following: Horizontal adding; a modified spot-pattern test;* digit-letter substitution; and anagrams. The test which showed a non-linear plot, particularly between the intermediate and final distributions, was a modification of Philip's multiple instruction cancellation test.

Following Thurstone, the test of linearity which has been used is a correlation near $+1.0$ in the x-values of overlapping distributions and the absence of systematic deviations from linearity in the plot of the pairs of their x-values. The size of the correlation between the x-values of overlapping distributions does not vary appreciably with change in the number of x-values used, when this number depends

*For a detailed description of this test, as well as of the other practice tests referred to, see (7).

merely on the class-interval in which the scores are grouped. For example, in one case the correlation remained exactly the same (+.991) for 9 pairs of x-values as for the 17 pairs of values resulting from the use of score-steps one-half as large. The correlation is apt to be affected, however, by the degree of overlapping of the distributions (the less the overlapping the more severe the test of linearity). Another factor which is likely to affect the correlation is the range of the x-values taken into consideration. Explicitly, the x-values computed from proportions near 1.0 or 0 are relatively unreliable, and if the x-values of scores constituting but one or two percent of a small total population are included, the correlations are apt to fall off. In the present study, only x-values for scores which were exceeded by at least 7.5 but not more than 92.5 percent of the cases were used for scaling. The eight correlations between the x-values represented by the two plots for each of the four tests subjected to scaling ranged from +.986 to +.999. Only one of the correlations was less than +.990, that of +.986 between the initial and intermediate distribution of horizontal adding scores. The correlations are given in detail in Table I.

TABLE I

Correlations of x-values of Adjacent Distributions

D_1, D_2, and D_3 stand for initial, intermediate and final score distributions. n stands for the number of x-values correlated.

Test	D_1 and D_2		D_2 and D_3	
	r	n	r	n
Horizontal Adding	+.986	12	+.990	16
Spot-Pattern	+.995	16	+.993	13
Anagrams	+.991	17	+.994	19
Substitution	+.996	17	+.999	21

These correlations, while very high, may nevertheless not suffice in all cases, particularly in view of the fact that the x-values for the extremes of the distributions were not included, to demonstrate a strictly linear relationship (see **2**). They do indicate, however, that the data are compatible with the assumption that the distribution of ability in the various cases was approximately normal.

While Thurstone's scaling procedure is presumably well-known, the application of the method to practice data is novel. The exact procedure used will therefore be illustrated by describing the steps followed in the case of anagrams. The first step is that of obtaining the distributions to be scaled. Distribution D_1 consists of the scores made

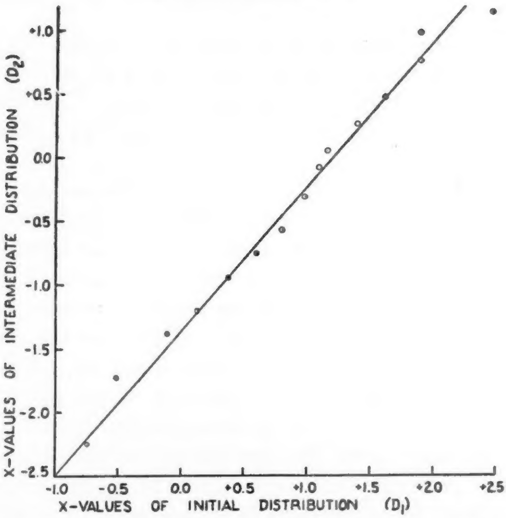


FIG. 1.—Plot of the x-values for all scores not exceeded by more than 99 percent or less than 1 percent of the score populations of D_1 and D_2, in the case of anagrams. The straight line in this figure, as also in Figure 2, is the line of best fit for those x-values corresponding to percents between 7.5 and 92.5, the only x-values used in scaling.

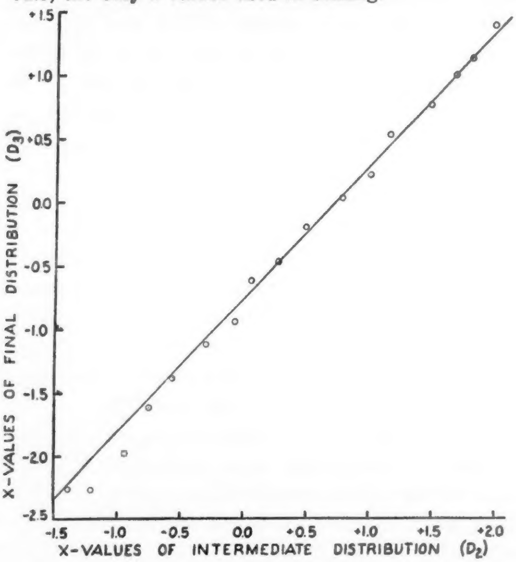


FIG. 2.—Plot of the x-values for distributions D_2 and D_3, in the case of anagrams.

on the first five days of practice. The mean scores for each of these days in order from the first, were: 30.4, 32.3, 32.6, 34.7, and 40.3. As already stated the individual scores for any one of these days was multiplied by the mean for the five days, divided by the mean of the day in question. For example, each of the scores made on the first day was multiplied by 34.1/30.4 or 1.122. In this way the raw scores were transformed so that the mean for each of the five days equalled the mean of the original scores for all five days. After the scores for each day were adjusted in the manner described, the scores for all five days were pooled. The same procedure was used to obtain the distribution D_2, from the scores for days 16, 17, and 18, and the distribution D_3, from the scores of the last three days, days 37, 38, and 39. Each pooled distribution was then tabulated in terms of cumulative percents so as to show the percent of scores falling below each raw score. In the case of other tests, the percentages calculated were those below every 2d, or every 5th score, when all possible scores were arranged in order. The percentages of scores falling below each raw score (extremes of distributions omitted) in the case of anagrams are shown in Table II.

The x-value of each percentage, based on the assumption of a normal distribution, was then obtained from the Kelley-Wood tables. In the present instance, the correlations of all pairs of x-values lying between the values for percentages of 7.5 and 92.5 were $+.991$, in the case of D_1 and D_2, and $+.994$, in the cast of D_2 and D_3.

The next step is to calculate the ratio of the σ's of the distributions D_1, D_2, D_3, in terms of the scaled scores, taking the σ of the distribution D_1 as unity. This is accomplished by formulae of the sort

$$\sigma_1/\sigma_2 = s_2/s_1$$

in which s_1 and s_2 represent the standard deviations of the x-values corresponding to a set of raw scores common to both D_1 and D_2. The values of s_1 and s_2, using only scores with a frequency not exceeding 92.5 or falling below 7.5 percent in either distribution D_1 or D_2, are .453 and .523; and for s'_2 and s_3 (from D_2 and D_3) are .609 and .642. It follows that the standard deviations of the three distributions, taking that of D_1 as unity, are 1.000, .866, and .821.

Since the x-values of each distribution are in terms of the σ of that distribution, the next step is to reduce all x-values to values in terms of the σ of the initial distribution, taken as unity. This is done by multiplying each x-value by the σ of the distribution to which it belongs (using the values 1.000, .866, and .821 in the case of D_1, D_2,

TABLE II

Distributions of Raw Scores in Anagrams
at Various Stages of Practice
The table shows the percentages of the scores which were lower than each
of the raw scores listed in the left-hand column.

Raw Score	D_1 (Days 1-3)	D_2 (Days 16-18)	D_3 (Days 37-39)
23.5	.075		
24.5	.136		
25.5	.189		
26.5	.228		
27.5	.271		
28.5	.304		
29.5	.396		
30.5	.457	.083	
31.5	.493	.083	
32.5	.550	.113	
33.5	.596	.137	
34.5	.646	.173	
35.5	.686	.190	
36.5	.725	.226	
37.5	.746	.250	.077
38.5	.786	.286	.083
39.5	.832	.345	.095
40.5	.836	.381	.131
41.5	.854	.423	.149
42.5	.861	.470	.173
43.5	.868	.488	.226
44.5	.875	.518	.268
45.5	.900	.554	.286
46.5	.918	.607	.321
47.5		.637	.369
48.5		.684	.423
49.5		.750	.476
50.5		.780	.512
51.5		.821	.548
52.5		.839	.589
53.5		.857	.643
54.5		.875	.702
55.5		.899	.744
56.5			.780
57.5			.810
58.5			.833
59.5			.857
60.5			.869
61.5			.899
62.5			.917

and D_3, respectively) and then adding to the resulting values the distances between the means of the distributions. The distance between the means of two adjoining distributions is determined by formulae of the sort

$$M_2 - M_1 = m_2 \sigma_2 - m_1 \sigma_1$$

in which M_2 and M_1 represent the scaled means and m_2 and m_1 the mean in distributions D_2 and D_1 of the x-values for identical scores, i.e., the scores whose x-values were used in calculating s_1 and s_2. In the case of anagrams, taking the mean of D_1 as an arbitrary zero, the mean of D_2 is $+1.197$ and of D_3, $(M_2 - M_1) + (M_3 - M'_2)$ or $+1.823$. By carrying out these operations, several scale values are obtained for all those scores common to several distributions. The extent to

TABL III

Scale Values of Raw Scores Calculated for Each Distribution

Raw Score	Scale Values D_1	D_2	D_3
23.5	—1.44		
24.5	—1.10		
25.5	— .88		
26.5	— .75		
27.5	— .61		
28.5	— .51		
29.5	— .26		
30.5	— .11	.00	
31.5	— .02	.00	
32.5	+ .13	+ .15	
33.5	+ .24	+ .25	
34.5	+ .37	+ .38	
35.5	+ .48	+ .44	
36.5	+ .60	+ .55	
37.5	+ .66	+ .61	+ .65
38.5	+ .79	+ .71	+ .69
39.5	+ .96	+ .85	+ .75
40.5	+ .98	+ .93	+ .90
41.5	+1.05	+1.03	+ .97
42.5	+1.09	+1.13	+1.05
43.5	+1.12	+1.17	+1.21
44.5	+1.15	+1.24	+1.32
45.5	+1.28	+1.31	+1.36
46.5	+1.39	+1.43	+1.44
47.5		+1.50	+1.55
48.5		+1.61	+1.66
49.5		+1.78	+1.77
50.5		+1.87	+1.85
51.5		+1.99	+1.92
52.5		+2.05	+2.01
53.5		+2.12	+2.12
54.5		+2.19	+2.26
55.5		+2.30	+2.36
56.5			+2.46
57.5			+2.54
58.5			+2.62
59.5			+2.70
60.5			+2.74
61.5			+2.87
62.5			+2.96

which the various scale values of the raw scores agree may be judged from Table III.

The final scale value for any raw score may be determined by averaging the several slightly discrepant scale values for that score obtained from the several distributions. Theoretically, a weighted average should be used; and the scale should then be smoothed, either by the method of a moving average or by the use of a formula expressing the relationship between raw scores and scaled values.* But for some purposes sufficiently approximate results may be obtained, without averaging at all, by using the scale values obtained for the separate distributions. In the present study, scale values of particular scores were desired to determine the approximate correlation between initial and final ability, and initial ability and gain. For this purpose, scale values given by the initial distribution were used for the initial scores and scale values given by the final distributions for final scores.

The initial scores are less reliable than the final scores and approximately equal reliability of initial and final scores may be obtained only by taking as each subject's score the mean of the scores he made on the pooled days. The σ of the mean scores, however, is inevitably smaller than the σ of the scores of any one day, or, what amounts to the same thing, the σ of the pooled scores.

There are, then, two questions: First, the effect of practice on the σ of equally reliable mean scores; and second, the effect of practice on the σ of the scores made with the duration of the test kept constant. In Table IV, summarizing the results obtained by scaling, the σ's of the initial and final mean scaled scores are represented by σ_I and σ_F, respectively, while the σ's of the initial and final pools are represented by σ_i and σ_f respectively. The ratio σ_F/σ_I offers an estimate of the effect of practice on individual differences when the initial and final measures are in terms of a constant unit and possess approximately equal reliability. A value on the part of this ratio of less than unity indicates a decrease, and one of more than unity an increase, in individual differences in ability as the result of practice. In calculating the correlations between initial and final score, and initial score and gain, only the mean scores have been used.

The values given in Table IV show that scaling resulted in uniformity among the four tests as regards the regression coefficient, b_{FI}, in that the value thereof was always less than unity, and also in

*For a formula of this sort, see **(5)**, p. 152.

TABLE IV

The Effect of Absolute Scaling
On Various Measures of the Results of Practice

		Anagrams	Substitution	Spot-Pattern	Horizontal Adding
σ_F/σ_I	(raw scores)	1.108	1.568	.570	1.876
	(scale values)	.871	1.249	1.235	.703
σ_f/σ_i	(scale values)	.821	1.189	1.206	.644
r_{IF}	(raw scores)	.818	.587	.590	.744
	(scale values)	.817	.619	.613	.718
b_{FI}	(raw scores)	.906	.920	.336	1.396
	(scale values)	.712	.773	.757	.505
r_{IG}	(raw scores)	— .146	— .062	— .822	+ .301
	(scale values)	— .498	— .225	— .242	— .711
Group Means	Initial (raw)	34.1	395.8	54.9	47.1
	" (scaled)	x	x	x	x
	Intermediate (raw)	44.6	440.4	26.3	64.3
	" (scaled)	$x+1.197$	$x+\ .571$	$x+1.862$	$x+1.164$
	Final (raw)	51.2	487.2	10.2	85.6
	" (scaled)	$x+1.823$	$x+1.044$	$x+3.580$	$x+2.035$

uniformity in the sign of the correlation between gain and initial ability, in that the sign of this correlation was always negative. These two values are of course closely related, since the correlation between initial ability and gain may be calculated from the formula*

$$r_{IG} = (b_{FI} - 1)\,(\sigma_I/\sigma_G),$$

in which r_{IG} is the correlation between initial scores and gains, and σ_G is the σ of the gains, i.e., the differences between final and initial scores. Inasmuch as the term (σ_I/σ_G) is always positive, it is apparent that when the regression coefficient is less than unity, the correlation between initial ability and gain must be negative. Absolute scal-

*Readily derived from the more familiar formula,

$$r_{IG} = \frac{\sigma_F\, r_{IF} - \sigma_I}{\sqrt{\sigma_F^2 + \sigma_I^2 - 2r_{IF}\,\sigma_I\,\sigma_F}}$$

by writing σ_G for the denominator and then dividing both numerator and denominator by σ_I. For this latter formula, see Ghiselli, E. F. and Kuznets, G., *J. Educ. Psychol.*, 1937, **28**, 238.

ing exerted no significant effect in the case of any of the tests on the correlation between initial and final score.

The important change resulting from scaling is in the standard deviation of the distributions, more particularly, in the magnitude σ_F/σ_I. Like the raw scores, the scaled scores show that individual differences may either increase or decrease with practice, depending upon the test. It is interesting to observe that the effect of scaling on the value σ_F/σ_I is quite different in the case of the spot-pattern and horizontal adding tests.

REFERENCES

1. ANASTASI, A., *Differential Psychology,* 1937, p. 168.
2. HOLZINGER, K. J., "Some Comments on Professor Thurstone's Method of Determining Scale Values of Test Items," *J. Educ. Psychol.,* 1928, **19**, pp. 112-117.
3. THURSTONE, L. L., "Scale Construction with Weighted Observations," *J. Educ. Psychol.,* 1928, **19**, p. 444.
4. WOODROW, H., "Absolute Scaling Applied to Data Yielded by the Method of Constant Stimuli,"*J. Exper. Psychol.,* 1937, **20**, pp. 1-28.
5. WOODROW, H., "Two Quantitative Laws Relating to Goodness of Performance," *Journ. of Psychol.,* 1937, **4**, pp. 139-150.
6. WOODROW, H., "The Interrelation of Conditions of Difficulty," I, *Journ. Gen. Psychol.,* 1937, **16**, pp. 83-102 and II, *ibid.,* pp. 103-130.
7. WOODROW, H.,"The Relation of Abilities to Improvement with Practice," *J. Educ. Psychol.,* 1938, (in press).
8. WOODROW, H., "The Effect of Practice on Groups of Different Initial Ability," *J. Educ. Psychol.,* 1938, (in press).

P

d
A
f
th
m
in
it
S
th
in
al
th
va
w
is
pe

pr
ca
of
of
ti
an

va
of
w

PSYCHOMETRIKA—VOL. 2, NO. 4
DECEMBER, 1937

ABILITY, MOTIVATION, AND SPEED

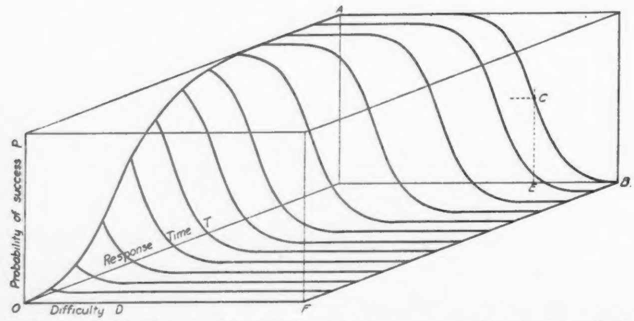
L. L. THURSTONE
The University of Chicago

The relations between ability, motivation, and speed in a mental task are represented by a psychometric surface. Ability of a subject is defined as the degree of difficulty for which the probability is 1/2 that he will complete the task in infinite time. Motivation affects the rate at which mental work is done, but not the altitude of performance. Suggestions towards experimental work are made.

In the appraisal of mental ability it is a recurring problem to distinguish between the profundity of a performance and its speed. A similar distinction is made between the relative difficulty of a performance and its speed. A high rating in an ability usually means the power to perform some kind of task that is beyond the attainments of most men. The performance of a difficult task is taken as indicative of ability more frequently than speed on simple tasks. Ability as power is a concept that Thorndike has denoted "altitude." Speed is the number of tasks that are completed in unit time, and these tasks are usually easy. This is the problem whether high speed in the performance of easy tasks can be taken as indicative of the ability to perform more difficult tasks without pressure of time. It is the latter form of ability that is usually regarded as the socially more valuable. The designing engineer, the architect, the scientist, the writer, are not evaluated in terms of stop watch performances. This is an old problem that has been the subject of quite a number of experimental studies.

It is the purpose of this paper to present a new approach to this problem. The analysis seems to show that mental ability as power can be experimentally determined so as to be independent of the speed of the performance. It also seems to be possible to appraise the ability of an individual so that the rating is independent also of his motivation. Our object is to appraise ability as power, independent of speed, and independent of motivation.

In the accompanying figure let the origin be at O. Let the three variables be 1) difficulty of task, D, as determined by the proportion of a standard group of subjects who perform the task successfully within a specified time limit; 2) the response time, T, for an individ-



ual subject; and 3) the probability, P, that the individual subject will successfully complete the task within a specified response time. The surface in the figure represents a single individual for a given kind of task and for a given degree of motivation. The task is sampled at different levels of difficulty, and each level is calibrated on a scale of difficulty D against a standard group of subjects. It is assumed that the individual subject attempts a large number of tasks at each level of difficulty D.

For any given response time T he solves a certain proportion P of tasks of difficulty D. This determines a point on the surface. If the difficulty D is set higher with the same response time, then the proportion P of successful solutions will decrease. This is represented by the fact that the surface slopes toward the right for increasing values of D. If the allowed response time T is increased without changing the difficulty D, then the proportion P of successful responses is increased. This is shown in the surface which rises with increasing values of T.

Consider the section PD at the origin. The surface cuts this section in the axis D. The ordinates P are identically zero for all values of D because the response time T is zero. This means that no performance of any difficulty can be completed in zero time. Parallel sections with small values of T show small proportions of the easier tasks completed but none of the more difficult.

Consider the section at AB with a large value of T. The surface cuts this section in the descending psychometric curve whose median point is at C. The corresponding median value of difficulty is shown at E on the base of the model. This is the level of difficulty for which the probability P is 1/2 that the subject can do the task. The more difficult the task, the lower is the probability that the task will be successfully completed in the time allowed. Hence, the psychometric curve is descending.

If the time allowed is already generous as represented by a large value of T, then an increment in T will have relatively little effect on the psychometric curve ACB in a plane parallel to PD. This well known fact suggests a definition of the ability of the individual subject with special reference to power or altitude which is independent of the speed of any performance. *The ability of an individual subject to perform a specified kind of task is the difficulty E at which the probability is 1/2 that he will do the task in infinite time.* In the illustrative diagram this ability is represented by E on the D-scale since E is the median for the psychometric curve ACB when T is infinite.

In the experimental determination of E it is not necessary to realize any very protracted time intervals. It is necessary only to determine experimentally a number of points on the surface for different finite values of time T and for different values of D. Since the effect on P of an increment in T becomes vanishingly small for large values of T, it is evident that the surface must become cylindrical and parallel to the axis T as T increases. Determination of E can be made by experimentally determining sections parallel to the plane of PT for several values of D. Each of these curves is a cumulative frequency distribution of response time. One of these integral frequency curves is shown in the plane of PT. Another such curve is shown in the line FB. It represents a level of difficulty F which is so high that the subject cannot do any of the tasks even in infinite time. Hence P is zero for all values of T when the difficulty is as high as F. The value of E can be found by interpolation. It is the value of D for the section, parallel to PT, whose cumulative frequency curve has an asymptotic limit of P that is 1/2. In this manner the absolute ability E can be experimentally determined even though it is defined as a performance in infinite time.

The origin for the scale of D should be such that the cumulative frequency distribution in the plane of PT represents simple reaction time without discrimination. The prototype for the psychological test is the psychophysical discrimination. All psychological test items can be regarded as extensions of psychophysical discrimination. Hence the psychometric curve ACB is fundamentally the same function as the curve of cumulative frequencies or percentiles that is in common use.

So far we have considered the appraisal of ability in the sense of power as distinguished from all measures that involve speed of performance. We have seen that the appraisal of ability as power can be made independently of speed of performance. Motivation is an-

other variable that is not directly measurable, but it has profound effects on any appraisal of ability. Let us assume, as an approximation, that the motivation of a subject to perform well remains essentially constant throughout the experiments with any one kind of task, although these tasks vary among themselves in difficulty. This assumption seems plausible. The surface could be determined experimentally on this assumption. Now suppose that the motivation were given a finite increment and that the surface were again determined experimentally. For finite time allowances we should expect higher values of P. In other words, if the subject can do a task at all, the effect of an increment in motivation should be to decrease his response time. The surface should move parallel to the axis T so as to augment the ordinates P for finite values of T.

If the subject fails a task in generous or infinite time at a certain level of difficulty, he may not be able to do the task even if his motivation were augmented. This is an interesting hypothesis. If the hypothesis is correct, it would lead to a fruitful theorem, namely, *that the ability E, determined experimentally as an interpolated asymptote, is independent of motivation if the motivation is itself greater than zero.* Another way of stating this hypothesis is that if motivation has any positive value, then its magnitude or intensity has no effect on the performance if infinite time is allowed for its completion. To say that motivation has a positive value is merely to say that the subject has some desire to do the task. If he were completely indifferent, he would make no attempt, and his motivation would be zero. Negative motivation would mean aversion or active refusal to do the task. A scale for measuring the intensity of motivation can be built on the definition of motivation as the first derivative of satisfaction[1] which is the psychophysical S-variable in the law of comparative judgment.[2]

The geometrical interpretation of this hypothesis is that several levels of motivation are represented by several surfaces. These surfaces are all identical except for displacements parallel to the axis T. If the surface in the figure is cylindrical and parallel to the axis T for large values of T, and if a change in motivation is represented by a displacement of the surface parallel to T, then it follows that the psychometric curve ACB and the absolute ability E are independent of motivation. A family of surfaces that represent different degrees of

[1]Thurstone, L. L., "The Indifference Function," *Journal of Social Psychology*, Volume II, No. 2, May, 1931.

[2]Thurstone, L. L., "The Law of Comparative Judgment," *Psychological Review*, Volume 34, No. 4, July, 1927.

motivation would become a single surface for large values of response time.

If this hypothesis is correct, it would follow that motivation is a rate concept. With reference to overt performance it would define only the rate of mental work. Motivation should have no effect on the question whether an individual is able to solve a problem. It should determine only how quickly he can solve it. According to this hypothesis we should not tell a man that he can solve a problem if he tries hard enough. We should tell him that if he can do the problem, then he can do it faster if he tries hard. This reasoning applies to problem solving and to discrimination of many kinds, but it cannot apply to a task which is by definition a rate of work.

Practice on the kind of task which is represented by the axis D may have the effect of improving the performance. Even the performance with generous response time may be improved with practice or instruction. Such improvement would be represented in the diagram by a displacement of the surface parallel to the axis D. If a discrimination can be improved with instruction or practice, the psychometric curve ACB will move toward the right and parallel to the axis D. The median score E will then also rise.

The general shape of the surface can be readily determined as shown here by well known psychological facts. This general shape is indicated by the lines which represent sections parallel to the section ACB. It is an interesting question of experimental fact whether these parallel sections are all identical psychometric curves except for displacement parallel to the axis D. It could readily be determined experimentally whether the standard deviations of these curves are all the same or whether the standard deviation of these sections is a monotonic function of T. This question could be answered by experimental data on the frequency distributions of response times for discriminations of pairs of stimuli whose separations represent difficulty D. These cumulative frequency distributions would determine the same surface with sections that are parallel to the plane PT. These questions of experimental fact may not affect the present hypothesis that the surface becomes cylindrical and parallel to the axis T for large values of T. It is a question of experimental fact whether an increment of motivation is represented by a translation of the surface parallel to the axis T. If this should be found to be the case, then the ability or power E is independent of motivation.

The present analysis has revealed relations between the difficulty of a task, the allowed response time, and motivation as independent

variables and the probability of successful completion of a specified kind of task as the dependent variable. The ability of an individual in a specified kind of task, in the sense of *power* which is independent of speed of performance, has been defined as that degree of difficulty for which the probability is 1/2 that he will complete the task in infinite time. It has been shown how this appraisal of power, independent of speed, can be made experimentally as an interpolated asymptote even though the definition implies infinite time which cannot be directly realized experimentally. Motivation may have no effect on the appraisal of power if the motivation of the subject has any positive value. Motivation can be interpreted as a rate concept which describes the rate at which mental work is done. According to this interpretation, the appraisal of an individual's ability for a specified kind of power task, as distinguished from tasks involving rate, can be made experimentally so as to be independent of speed of performance and also independent of his motivation.

PSYCHOMETRIKA—VOL. 2, NO. 4
DECEMBER, 1937

SOME ASPECTS OF RASHEVSKY'S THEORY OF DELAYED REFLEXES

A. S. HOUSEHOLDER AND E. AMELOTTI
University of Chicago

A closed solution of the integral equation obtained by N. Rashevsky, with the assumption that the inhibitory influence between centers is a constant, i.e., independent of the distance apart, is obtained. Furthermore, a more general kernel, representing a variable inhibitory influence, which in our case is a monotonic (increasing or decreasing) function of the distance between centers, is introduced. The resulting integral equation is solved and some properties of the solution discussed.

In a series of papers (**1**) N. Rashevsky has proposed a theory of delayed reflexes somewhat as follows: A certain stimulus S_u arouses nervous impulses leading to a given response R_u, which impulses spread out and must pass thru a particular group of synapses Σ before reaching the motor nerves which ultimately effect the response R_u. These synapses Σ are arranged in linear sequence and may be thought of as points ξ on a segment of a straight line, say from $\xi = 0$ to $\xi = \xi_0$. Moreover the excitation from S_u reaches all these synapses simultaneously.

These synapses Σ are also reached by another set of afferent neurones which may be aroused by a stimulus S_c, the conditioned stimulus. However, initially the stimulus S_c is not adequate actually to produce R_u. Furthermore, the excitation from S_c does not reach all centers simultaneously, but reaches first the center ξ_0 and then passes on to the others in order with a finite constant velocity. Thus at a time, say, τ seconds after the excitation reaches ξ_0, it has reached a further center ξ_1, and, as the excitation persists, though with diminishing intensity, all centers from $\xi = \xi_1$ to $\xi = \xi_0$ are excited at this time to an extent which decreases with increasing ξ. The expression assumed for this excitation curve is

$$I(x) = I_0 e^{-\beta x}, \quad x = \xi - \xi_1. \tag{1}$$

Now if at this time τ the centers are excited from the stimulus S_u, so that R_u is actually produced, then there results a lowering of the threshold to S_c, and in fact with n repetitions of this superposition

of the two types of excitation, the capacity of S_c to excite the efferent neurones from Σ becomes

$$R = F\,I(x)\,(1 - e^{-an}), \tag{2}$$

where F and a are positive parameters. It is understood that in every case the unconditioned stimulus is not applied until τ seconds after the conditioned.

Finally suppose that the efferent neurone from each center x is connected by an inhibitory fiber to every other center s. Then the excitation is no longer the applied excitation (1), but is given as the solution of an integral equation of the following form:

$$I(x) = I_0\,e^{-\beta x} - \lambda \int_0^k K(x,s)\,I(s)\,ds. \tag{3}$$

In this equation the upper limit k is taken as $x_0 = \xi_1 - \xi_0$ unless the $I(x)$ vanishes for a smaller value $x = k$. The kernel K measures the inhibitory effect of center s upon center x, and is always positive or zero. The parameter λ is

$$\lambda = b\,F\,(1 - e^{-an}), \tag{4}$$

where b is another parameter. Rashevsky considered only the case of a constant inhibitory effect with $K \equiv 1$, but otherwise the derivation of equation (3) is the same as the derivation of his equation II(8).

If $I(x)$ does vanish for some $k < x_0$, then from equation (2) we see that when conditioning is complete and R_u does follow S_c, it can do so only after the excitation has time to pass from center x_0 to center k. Moreover, this point k may evidently vary with λ, which in turn varies with n. Hence just adequate conditioning may well give one value of k, while more complete conditioning might lead to a different k and hence to a different time-interval between the application of S_c and the actual production of R_u. In particular it may happen that k diminishes and the time-interval therefore increases. We are here concerned with the existence and variation of k with λ for some special types of kernel K, which yield a closed solution to the integral equation (3).

We note here that in general the solution I is a function of k and of λ as well as of x:

$$I = I(x, k, \lambda).$$

But k is determined as the smallest positive zero of the equation

$$I(k, k, \lambda) = 0 \tag{5}$$

and this gives k as a function of λ. Now if such a k can be found, then $I(x)$ has the same sign on the interval from 0 to k, since otherwise a smaller k would have been chosen. Moreover, on this interval, $I(x)$ is necessarily everywhere positive, since if it were everywhere negative the right member of (3) would be positive, which is a contradiction.

If K can be factored

$$K(x, s) = \varphi(x)\, \psi(s) \tag{6}$$

as the product of a function of x alone and a function of s alone (2), a closed solution can always be found by quadratures. For, consider the equation

$$I(x) = f(x) - \lambda\, \varphi(x) \int_{k_1}^{k_2} \psi(s)\, I(s)\, ds. \tag{7}$$

Then we may write

$$I(x) = f(x) - \lambda\, \mu\, \varphi(x), \tag{8}$$

where

$$\mu = \int_{k_1}^{k_2} \psi(s)\, I(s)\, ds. \tag{9}$$

Hence substituting (8) into (9) and solving for μ,

$$\mu = \frac{\int_{k_1}^{k_2} \psi(s)\, f(s)\, ds}{1 + \lambda \int_{k_1}^{k_2} \psi(s)\, \varphi(s)\, ds}, \tag{10}$$

and this, substituted back into (8) gives the solution. In the special case,

$$\varphi = \psi = 1,\ f(x) = I_0\, e^{-\beta x}, k_1 = 0,\ k_2 = k, \tag{11}$$

and hence

$$\mu = I_0 (1 - e^{-\beta k}) / [\beta (1 + \lambda\, k)]. \tag{12}$$

This gives, for Rashevsky's integral equation, the solution

$$I(x) = I_0 [e^{-\beta x} - \lambda (1 - e^{-\beta k}) / \{\beta (1 + \lambda\, k)\}]. \tag{13}$$

To find k, we replace x by k and set the result equal to zero. After a little algebraic manipulation we obtain the equation

$$e^{\beta k} - \beta k - 1 = \frac{\beta}{\lambda}. \tag{14}$$

The left member of this equation in k vanishes with k and increases monotonically. It is, moreover, independent of λ. The right member is always positive, and decreases as λ increases. Hence for every λ, and therefore for every n, equation (14) is satisfied by a unique value of k, and this value decreases monotonically as λ, and hence as n, increases. For any such value of k the corresponding solution to the integral equation may be written more simply as

$$I(x) = I_0(e^{-\beta x} - e^{-\beta k}). \tag{15}$$

Thus we find further that as n increases and k decreases, $I(x)$ decreases at each center x.

I. *In the case of constant inhibitory influence, if x_0 is sufficiently large the centers of Σ last reached by the excitation from S_c completely blot out the excitation of those first excited, and the extent of the interval so blotted out increases as the conditioning becomes more complete. The center k which is the first of the blotted out centers is given as the unique solution to equation* (14), *and the excitation $I(x)$ is then given by* (15), *decreasing for every x as the conditioning proceeds. Finally, k approaches asymptotically a value obtained by replacing λ by bF in equation* (14).

It is natural to inquire next about the effect of a variable inhibitory influence which is a monotonic (increasing or decreasing) function of the distance between the centers x and s. We therefore introduce the kernel

$$K(x, s) = e^{m|x-s|}, \tag{16}$$

and consider the resulting equation

$$I(x) = I_0 e^{-\beta x} - \lambda\, e^{mx} \int_0^x e^{-ms}\, I(s)\, ds - \lambda\, e^{-mx} \int_x^k e^{ms}\, I(s)\, ds. \tag{17}$$

The two integrals are needed because of the fact that the absolute value of $x - s$ is to be used in the exponent. Here again a closed solution is obtainable. Without carrying out the procedure in general, it may be pointed out that the method of solving depends only upon possibility of factoring $K(x, s)$ on each of the two intervals $(0, x)$ and (x, k) into functions of x alone and of s alone.

In our case we set

$$\varphi(x) = e^{mx} \int_0^x e^{-ms} I(s)\,ds,$$
$$\psi(x) = e^{-mx} \int_x^k e^{ms} I(s)\,ds, \tag{18}$$

whence

$$I(x) = I_0\, e^{-\beta x} - \lambda(\varphi(x) + \psi(x)). \tag{19}$$

Dividing thru by the exponential in each of equations (18), differentiating, and again dividing out exponentials, we obtain the system of differential equations for $\varphi(x)$ and $\psi(x)$:

$$(D - m)\varphi = I(x) = I_0\, e^{-\beta x} - \lambda(\varphi + \psi),$$
$$(D + m)\psi = -I(x) = -I_0\, e^{-\beta x} + \lambda(\varphi + \psi), \tag{20}$$

where

$$D = \frac{d}{dx}. \tag{21}$$

From these two equations we eliminate ψ and obtain

$$(D^2 - m^2 + 2\,m\,\lambda)\varphi = I_0(m - \beta)\, e^{-\beta x}. \tag{22}$$

We consider only the case

$$m^2 - 2\,m\,\lambda > 0,$$

which certainly holds if

$$m < 0 \quad \text{or} \quad m > 2\,b\,F.$$

With this restriction we may set

$$\Delta^2 = m^2 - 2\,m\,\lambda, \qquad \lambda = (m^2 - \Delta^2)/2\,m. \tag{23}$$

Then the solution of (22) is

$$\varphi(x) = [I_0(m-\beta)/(\beta^2 - \Delta^2)]\,[e^{-\beta x} - c_1\, e^{\Delta x} - c_2\, e^{-\Delta x}], \tag{24}$$

whence the first of equations (20) gives

$$I(x) = [I_0(\beta-m)/(\beta^2 - \Delta^2)]\,[(\beta+m)\, e^{-\beta x}$$
$$+ c_1(\Delta-m)\, e^{\Delta x} - c_2(\Delta+m)\, e^{-\Delta x}]. \tag{25}$$

The constants c_1, c_2, and k must then be determined from the set

$$\varphi(0) = 0, \qquad \psi(k) = 0, \qquad I(k) = 0, \tag{26}$$

or the equivalent set,

$$\varphi(0) = 0, \qquad \lambda\,\varphi(k) = I_0\, e^{-\beta k}, \qquad I(k) = 0. \tag{27}$$

If we solve the last two of these equations for c_1 and c_2 and then substitute into the first we obtain, after some algebraic manipulation, the following equation for determining k:

$$\left.\begin{aligned} &e^{-(\beta+\Delta)k}(\Delta-\beta)(\Delta+m)/(\Delta-m) + e^{-(\beta-\Delta)k}(\Delta+\beta) \\ &\quad\times(\Delta-m)/(\Delta+m) - 2\,\Delta(\beta-m)/(\beta+m) = 0. \end{aligned}\right\} \tag{28}$$

In discussing the existence of solutions k of this equation it is convenient to represent the three terms in this equation by $f(k)$, $g(k)$ and $-\mu$, respectively. Hence we may write for (28)

$$F(k) \equiv f(k) + g(k) - \mu = 0. \tag{29}$$

By elementary calculations we find that

$$F(0) = 4\,\Delta\, m(\Delta^2 - \beta^2)/[(\Delta^2 - m^2)(\beta + m)]. \tag{30}$$

Moreover we find that if

$$\kappa = \frac{1}{\Delta}\log\frac{m+\Delta}{m-\Delta} \qquad (m > 0), \tag{31}$$

then

$$\left.\begin{aligned} F'(\kappa) &= 0 \qquad (m > 0), \\ F'(k) &\neq 0 \qquad (m < 0). \end{aligned}\right. \tag{32}$$

That is, F' vanishes only at κ if $m > 0$, and F' vanishes nowhere if $m < 0$. It is easy now to verify the following tabulations:

		$F(0)$	$F(\kappa)$	$F(\infty)$	$F'(0)$	$F'(\infty)$
$m > 0$	$\beta > m > \Delta$	+	—	—	+	—
	$m > \beta > \Delta$	+	—	+	—	+
	$m > \Delta > \beta$	—	+	$-\infty$	+	—
$m < 0$	$\beta > \Delta > -m$	+		—	—	—
	$\Delta > \beta > -m$	—		$+\infty$	+	+
	$\Delta > -m > \beta$	+		$+\infty$	+	+

We conclude, therefore, that if x_0 is sufficiently large, then there always exists a solution $k < \kappa$, when $m > 0$, while if $m < 0$ the solution k exists provided only $\beta > -m$.

To examine the concentration, observe that Δ is monotonic in λ and hence in n, decreasing for $m > 0$, increasing for $m < 0$. Consequently we need only examine the derivative $dk/d\Delta$ of k as defined by equation (29) as a function of Δ. But since

$$F_k \frac{dk}{d\Delta} + F_\Delta = 0 \tag{33}$$

we may compare the signs of F_k and F_Δ. Hence we consider

$$\begin{aligned} f_\Delta &= \left[\frac{(\Delta - m)^2 + 2m(\beta - m)}{(m^2 - \Delta^2)(\beta - \Delta)} - k\right] f, \\ g_\Delta &= \left[-\frac{(\Delta + m)^2 + 2m(\beta - m)}{(m^2 - \Delta^2)(\beta + \Delta)} + k\right] g, \\ \mu_\Delta &= 2(\beta - m)/(\beta + m), \end{aligned} \tag{34}$$

and show that for sufficiently large β, or for sufficiently small m and Δ, concentration always takes place.

First consider the effect of increasing β. For any fixed k, f and g may be made as small as desired by choosing β sufficiently large. Hence for any k, $F(k)$ can be given the sign of $-\mu$, which is the sign of $F(\infty)$, by choosing β sufficiently large. Hence the root k of (29) is bounded as β increases, and becomes, in fact, vanishingly small. Hence f_Δ and g_Δ can be made as small as desired, so that F_Δ acquires the sign of $-\mu_\Delta$. But this is opposite that of $F'(k)$ for $m > 0$, and the same for $m < 0$. Hence $dk/d\Delta$ is positive for $m > 0$, negative for $m < 0$, and in either case k is monotonically decreasing in λ and in n.

For given β and small m and Δ, we may make the replacements

$$\beta = \pm m\gamma, \qquad \Delta = \pm m\delta, \quad \pm mk = \mu, \tag{35}$$

according as m is positive or negative. Decreasing m has the effect of increasing γ, while Δ may be kept fixed by decreasing bF. The same argument as the preceding then applies.

II. *When the mutual inhibitory effect of the centers is given by the kernel* (16), *the "blotting out" and the concentration effects are the same as for a constant kernel provided the parameter β is sufficiently large or provided the parameters m and bF are sufficiently small. This is true moreover, whether m is positive or negative.*

As seen from equation (22), two possible cases arise depending upon the sign of $m^2 - 2\lambda m$. We have here considered the cases arising when this is positive. If, however, it is negative, then represent-

ing this by $-\Delta^2$, we have the solution $I(x)$ involving $\sin \Delta x$, $\cos \Delta x$, instead of the exponentials which we have considered. Hence added to the monotonically decreasing term $e^{-\beta x}$ is an oscillating function with period $\frac{2\pi}{\Delta}$.

Interesting possibilities are presented by this form of the function. It may be that there is a unique k as before, and if so the concentration or lack of it needs be considered. However, it may happen that the excitation $I(x)$ dies out completely in a series of intervals between which it remains greater than zero. If so, in the successive intervals within which $I(x)$ is greater than zero, $I(x)$ takes on maxima, which diminish successively due to the presence of the decreasing exponential term.

In concrete terms, consider the conditioning of an animal to salivation at the sound of a bell. Instead of salivation 20 seconds, say, after the sound of the bell, we should find, perhaps, at intervals of 10, 15, 20, seconds that salivation recurs at increasing degree.

Whether this could be so depends upon the relations between the parameters involved in the integral equation, i.e., upon the properties of the nerve centers involved. In any case it should be pointed out that the trigonometric functions cannot arise at all unless $0 < m < 2\lambda$, and this is not true unless $m < 2bF$. The condition $m > 0$ means that the inhibitory effect between centers far apart is greater than that between centers close together; and the second condition that $m < 2bF$ limits the amount of the variation of the effect.

REFERENCES

1. N. RASHEVSKY, "Mathematical Biophysics and Psychology," *Psychometrika*, 1936, **1**, pp. 1-26.
   N. RASHEVSKY, "Mathematical Biophysics of Delayed Reflexes in Connection with the Theory of Error Elimination." *Psychometrika*, 1936, **1**, 265-273. We shall refer to this paper as II.
2. In fact if K is expressible as a sum of finite number of such products. See, *e. g.*, Courant-Hilbert, *Methoden der Mathematischen Physik*, pp. 99-101, 1931 (Berlin).

PSYCHOMETRIKA—VOL. 2, NO. 4
DECEMBER, 1937

A FACTOR ANALYSIS OF CERTAIN NEUROTIC SYMPTOMS

CHARLES I. MOSIER
University of Florida, Gainesville, Florida

In an attempt to investigate the concept of neurotic tendency thirty-nine of the forty-two most discriminative items in the Thurstone Neurotic Inventory were administered as a questionnaire to a group of five hundred male college students. An analysis of the table of intercorrelations by Thurstone's centroid method showed that eight factors were sufficient to account for the observed intercorrelations with negligible residuals. The eight centroid factors were then transformed into a simple structure. It is concluded that a single trait of neurotic tendency cannot be postulated and tentative hypotheses are formed as to the nature of the primary traits revealed by the analysis.

Readministering the same test to the same students a week later showed a high consistency of response both on the test as a whole, and on the individual items.

The purpose of this investigation is to apply the methods of multiple-factor analysis to certain questionnaire responses of the type generally termed 'neurotic symptoms,' with a view to determining whether there is a single factor of neuroticism common to all of them, and what behavior traits must be postulated to account for the observed interrelations of these items of behavior.

The writer has elsewhere (**32**) inquired into the validity of the present questionnaire type of personality test and the traits there measured. The present investigation is a consequence of the conclusions there reached. It is designed to go back of the total scores on personality questionnaires of the type represented by the Thurstone Neurotic Inventory (**44**), the Bernreuter Scales (**4**) and Flanagan revision (**13**), the Woodworth Personal Data Sheet (**14**), and others of a similar nature, in order to test the hypothesis that the scales are actually measuring a primary trait of neurotic tendency, and if they are not, to lay the basis for further work in the development of scales that will measure such single, psychologically significant traits.

The method of this present investigation is broadly as follows: the forty-two most diagnostic items from the Thurstone Neurotic Inventory (**44**) were administered as a questionnaire to a group of college men. For thirty-nine of the items, the intercorrelations between each item and every other item, and between each item and the A. C.

E. Psychological Examination were computed, and the resulting table of intercorrelations subjected to a multiple factor analysis by the centroid method of Thurstone (**42**). The resulting factorial matrix was then transformed into a simple structure (**42**, ch. VI) to determine the primary traits underlying this set of neurotic symptoms.

The concept of primary "trait" as used in this discussion (**42**, p. 51) is an operational definition, involving the operations of correlation, factor analysis and transformation to a simple structure. The significance of such a concept may best be followed in Thurstone's *Vectors of Mind*, pp. 44-54 (**42**). From there is quoted the following:

> It is the faith of all science that an unlimited number of phenomena can be comprehended in terms of a limited number of concepts, or ideal constructs. . . The constructs in terms of which natural phenomena are comprehended are man-made inventions. To discover a scientific law is merely to discover that a man-made scheme serves to unify, and thereby to simplify, comprehension of a certain class of natural phenomena. A scientific law is not to be thought of as having an independent existence which some scientist is fortunate to stumble upon. A scientific law is not a part of nature. It is only a way of comprehending nature. . . Just as it is convenient to postulate physical forces in describing the movements of physical objects, so it is also natural to postulate abilities [or traits] and their absence as primary causes of the successful completion of a task by some individuals and of the failure of other individuals in the same task.
>
> The criterion by which a new ideal construct in science is accepted or rejected is the degree to which it facilitates the comprehension of a class of phenomena which can be thought of as examples of a single construct, rather than as individualized events. . . It is in the nature of science that no scientific law can be proved to be right. It can only be shown to be plausible. The laws of science are not immutable. They are only human efforts toward parsimony in the comprehension of nature.

The particular items chosen for this investigation were selected for two reasons: first, because the method of their determination by the authors of the Neurotic Inventory (**40**) indicated that if there were a common trait of emotionality or neuroticism, these items should manifest it; and second, they are representative of the current scales and measures as indicated in the following analysis, in which is shown for each of fifteen scales or studies the number of items identical with items in the thirty-nine finally selected.

Bernreuter-Flanagan (**4**, **12**)	29
Bell Adjustment Inventory (**3**)	24
Harvey (**18**)	23
Willoughby (**46**)	18
Landis Inventory (**24**)	15
Laird B-2 (**23**)	13

Page Schizoid Traits (**33**)	12
Woodworth P. D. Sheet (**14**)	12
Maller Character Sketches (**28**)	12
Freyd-Heidbreder I-E Scale (**19**)	8
Guilford (**16**)	7

The items as presented to the subjects are reproduced in Table I. Three items, (nos. 14, 17, and 37) of the forty-two originally selected by the Thurstones as the most discriminating were rejected from this study after administration of the questionnaire because of the consistently low correlations of these items with the remaining items and with each other. The subject's score on the A. C. E. Psychological Examination (total score) was recorded as "above the median", or "below the median", and treated as the fortieth item (no. 43).

It must be urged that the variables under investigation are not the presence or absence of certain objectively verifiable attributes, but rather items of behavior in the strictest sense—making a mark in one position rather than in another position when confronted with a printed verbal statement. The primary datum is not, for example, the answer to the question, "Does this man have stage fright?", but to the question, "Does this man *say* that he has stage fright?" This does not mean that it is not legitimate to generalize beyond that restricted interpretation, but it does mean that such generalization must be made with full realization of the exact nature of the data on which that generalization is based. Guilford and Guilford emphasize this when they say of their similar material (**17**):

> . . . . it is necessary to keep in mind the nature of the original data. We must constantly remember that the response of a subject may not represent exactly what the question implies in its most obvious meaning . . . responses represent the cooperative efforts of the subjects to give fair appraisals . . . of their behavior as they know it . . . The most parsimonious interpretation that can be given to the data is that we have applied 36 stimuli to 930 subjects . . . who by common instruction and in a common classroom situation have reacted to those stimuli in one of two ways. The intercorrelations between items then represent indicators of similar reactions to pairs of stimuli.

In order to give an example of the difficulties that may arise, the results may be anticipated by mentioning that item 24—"Are you touchy on various subjects?"—following as it does the question "Have you ever been depressed by low marks in school?", gives consistent evidence of having been interpreted by the college students to mean: "Are you touchy about various *college courses*?" The studies of Wrightstone

(48) and Landis and Katz **(25)** indicate that the percentage of agreement between questionnaire response and external criteria will, for such items as these, lie between 85% and 90%, so that generalization beyond the primary data is not impossible.

The subjects of this study were 500 male students of the University of Florida, gathered from classes in psychology, speech, political science, and English. For the most part they were freshmen or sophomores, and had never taken a course in psychology, or were taking their first course. The administration of the questionnaire was preceded by a short talk, urging the subjects' cooperation in a research project in psychology, stressing that the results were for research purposes only, and would be kept entirely confidential. The instructions were given orally, the effort being made to be sure that every subject understood what he was to do, rather than to keep the verbal form of the instructions rigidly uniform. Before the completed questionnaires were collected the subjects were told that if any of them, after seeing what was required of him, preferred not to turn in his paper, but to destroy it, that was permissible. No student availed himself of this privilege. Approximately one hundred papers were eliminated because of excessive number of "?" responses.

Without any warning, the questionnaire was administered a second time, as near to a week later as class schedules would permit. The results of this second administration were utilized only to obtain an estimate of the consistency with which each item was answered, and all subsequent computations except these "consistency coefficients" were based on the results of the first administration.

The items were scored according to the key furnished by the authors of the Neurotic Inventory, as neurotic or non-neurotic. In every case the "?" response was scored by the key as non-neurotic. In order to reduce the influence of the "?" response as a disturbing factor in the results, no paper was considered which returned more than three such responses. Five hundred cases were secured with both administrations of the questionnaire and with no more than three "?" responses. These cases formed the group on which this study was made.

Of the items selected for final analysis, only two were phrased negatively, so that the answer 'No' was considered neurotic. Because of the desirability of keeping the questions in their original form for purposes of comparison with other studies the wording of the items was left as it stood in the *Neurotic Inventory*. What effect this might have on the results is not known. Smith **(38)**, however, investigating the effect of positive and negative phrasing, finds that the positive phrasing gives more reliable results.

The frequency of neurotic response to each of the thirty-nine items is reported in Table II. These results agree closely with the item-frequencies reported by previous investigators.

For each of the 780 pairs of items on the first administration, and for the two administrations of each item, four-fold distribution tables were prepared, and the tetrachoric coefficients of correlation computed as described in Chesire, Saffir and Thurstone (**9**). The use of the tetrachoric coefficient assumes a normal correlation surface of two continuous variates. That this assumption is warranted is indicated by the distributions published by Laird (**20**) for similar items. The mean correlation, exclusive of self-correlations, is .198. The standard error of a tetrachoric coefficient of .198 is .0728 as computed by the formula given by Kelley, (**22**, p. 258).*

The matrix of intercorrelations was factored by the method presented in detail in *Vectors of Mind,* Appendix I (**42**), using as an estimate of the communality the highest value in the column and revising this estimate for each factor. The method of reflecting tests described in Appendix I was followed for the first five factors, after which it was deemed advisable to use the more accurate method of reflecting tests until the sum of each column, without the diagonal entry, was positive. The use of this method insures that each successive factor makes a maximum contribution to the total variance.

The factorial analysis was continued until eight reference factors had been determined, when the standard deviation of the frequency distribution of the 780 residuals was .057 and the maximum contribution of the eighth factor to any single correlation was .082. While further factors might have been extracted, it was felt that their contribution to the total variance would be negligible.† The factorial matrix of the forty items referred to the centroid reference factors is presented in Table III.

The factorial matrix Table III is a complete description of the observed interrelations between this set of forty items of behavior (except, of course, for the neglected contribution of the eighth factor residuals). It would be quite possible to postulate a 'trait' for each of the eight reference vectors represented by the columns of the factorial matrix of Table III. A description of the data in terms of such 'traits' would be a valid description, in that the observed intercorrelations could be reproduced by matrix multiplication of that matrix

*The complete matrix of intercorrelations is presented in Mosier, C. I., *A factor analysis of certain neurotic symptoms.* Ph.D. Thesis, Univ. of Chicago Library, 1937.

†The matrix of eighth-factor residuals is presented in Mosier, *loc. cit.*

by its transpose. Such a description, and such "ideal constructs" as those "traits" would be, have, however, certain disadvantages, viewed as a scientific description. First of all, there exist an infinite number of such descriptions, each involving only eight traits, and each equally valid. Furthermore, such a description would have to be revised with every alteration in the set of items investigated. (Whence arise the criticisms of factorial methods of Line, Griffin and Anderson (**26**), Russell (**35**), Burks (**6**), Allport (**1**), and Chant (**8**).) Finally, such a description is not the most parsimonious possible. To overcome these disadvantages, it is necessary to transform the centroid matrix to a simple structure (**42**, ch. vi), so that each item is explained in terms of the fewest possible number of the eight traits.

Lotka (**27**, p. 372) in an entirely different context summarizes the significance of 'simple structure' in the following quotation:

> Now an infinite number of equally appropriate pictures of this kind can be given for . . . the point, for the coordinates x, y, z, can be chosen in an infinite number of ways. But each choice of a system of coordinates will furnish its own particular set of functions 'f'; each picture will in this respect differ from every other such picture. Not a little of the success of the scientific investigator depends on his judgment in choosing a suitable system of coordinates, such as will furnish the simplest and most *convenient* picture, the most convenient functions of f. So, for example, the motion of the planets is most simply described with reference to a frame of coordinates fixed to the sun, and not, for example, one fixed to the earth. The discovery of a particularly convenient reference frame, though in a sense it adds nothing to our knowledge of concrete facts, may nevertheless constitute one of the major events in the progress of human knowledge.

In determining the simple configuration, the limitation that the primary traits be orthogonal was imposed after the oblique simple structure had been determined, and it had been seen that such a restriction would produce no great distortion from the oblique configuration. The assumption of orthogonality postulates traits uncorrelated in the population of individuals. It does not imply, as Allport seems to assume (**2**) that the traits are independent of one another in any but the statistical sense above. They may still "overlap", and may show interrelationships in the behavior items which are symptomatic of the traits.

The solution of the problem of simple structure was achieved by use of the procedure described by Thurstone (**43**), followed by graphic rotations until the simplest orthogonal configuration was attained. No assumption was made as to a positive manifold. The matrix of transformation L is reproduced in Table IV, and the Primary Trait

Matrix is presented in Table V. The significance of the identifying letters at the head of each column will be considered in the interpretations of the results. A sample of the graphic plots of each trait against every other for the final set of primary traits is presented in Fig. 1 showing traits '*C*' and '*S*'.

Table VI gives for each item its consistency coefficient—the self-correlation after an interval of one week—its communality and its specificity (**42**, p. 68). The consistency coefficients are comparable with those reported by Guilford and Guilford (**17**) for similar items. These are consistently higher than Guilfords', as would be expected from the fact that in the latter case the test-retest interval was one month. These consistency coefficients are subject to two distorting influences which prevent identification of them with the test-retest reliability coefficient, to which they bear formal correspondence. The first is the possibility that the subject remembered on the second occasion how he had answered on the first, and a desire to be consistent influenced the correlations. The second is that the answers were influenced by the mood of the moment, causing a lowering of the coefficients. The results of Johnson (**21**) bear on this second possibility.

The range of communalities is from .850 for item 41 to .153 for item 38. The specificities range from .009 for item 41, indicating that all of its variance not due to chance is explained by the eight primary traits, to a number of values greater than .50, indicating that for those items, the present investigation has accounted for less than half of the non-chance variance.

Interpretation of the primary traits—

Since the evidence upon which an interpretation of the psychological significance of the primary traits is based consists merely of the correlations obtaining between the items and that trait, the problem may be formulated thus: "What is the nature of the trait which would correlate x with item a, y with item b, and z with item c?" Rephrased, the problem might be stated: "What is the nature of the trait such that an individual who possessed it to a strong degree would tend to make responses a, b, and c, and tend not to make responses d, e, and f (taking into account the varying strengths of the tendencies to respond)?" Obviously the answer to such a problem cannot be given categorically. The tentative answers here advanced are presented not so much as answers or as definition of the traits, as in the nature of hypotheses, each of which requires independent verification.

In discussing each trait the items which have high loadings

("high" arbitrarily defined as .24 or over) will be presented separately for convenience. For the exact wording of the item, refer to Table I. For the first trait (column C in Table V) those items are:

41.	Ups and downs in mood	.85
11.	Happiness to sadness	.81
31.	Frequently in low spirits	.51
28.	Often just miserable	.43
10.	Worry over possible misfortune	.41
26.	Frequently feel grouchy	.39
40.	Cannot make up mind	.32
9.	Burdened by a sense of remorse	.30
25.	Often in a state of excitement	.29
16.	Interests change quickly	.29
22.	Mind wanders badly	.26
29.	Bothered by some useless thought	.25
3.	Worry over humiliating experiences	.25
6.	Feelings easily hurt	.25

Inspection of these items indicates that the underlying trait might be termed *'cycloid'*, indicative of the variability of mood expressed by the two highest items. The hypothesis of 'depression' must be rejected, since the second factor is clearly that. The close similarity between this trait and the trait 'E' reported by the Guilfords is all the more striking in view of the fact that only seven of their thirty-six items were similar to items in the present battery. In the trait as here determined, however, the cyclic nature of the trait is more apparent. This trait also bears similarity to Darrow's (**10**) constellation 'G'—excitability. The physiological measures of Darrow most consistently associated with this cycloid trait 'C' were his measures 9 and 18 (blood pressure rises after shock and percentage of association of the galvanic reactions to the conditioned stimuli with the conditioning stimuli). Both of these were negatively correlated with the items showing high loadings on 'C'. The blood pressure rise was most characteristically associated with the items having the lower loadings (items 25, 16 and 29).

The trait 'C' shows an interesting general correspondence to the personality items which Studman (**39**) finds characteristic of *'f'*. The correspondence between 'C' and Studman's *'f'*, particularly the objective tests of *'f'*, should be investigated. This trait shows similarities to the manic trait described by Thurstone (**41**) in his analysis of Moore's list of psychotic symptoms (**29**).

A possibility that should be considered in further investigations of the nature of this trait is that there exist two distinct traits, one manic in character and one depressive. The few manic and euphoric items in the present battery might not have been sufficient to separate the manic trait from a composite cycloid trait. The similarity between this trait and those determined by the Guilfords with a similar, and by Darrow with a different technique, raises the question whether it should not be called, as they have called it, "Emotionality". Since, however, emotionality is a rather vague term, and since the alternation of mood is the outstanding item in this and Guilfords' trait, it was felt that the term 'cycloid' described the most tenable hypothesis. To avoid crystallization of the hypothesis, however, the initial 'C' is to be preferred to the trait-name.

A consistent picture of *depression* is given by the second trait, designated 'D'. Its loadings are:

32.	Periods of loneliness	.71
4.	Lonesome, even with others	.68
31.	Frequently in low spirits	.59
28.	Often just miserable	.56
20.	Difficulty making friends	.50
26.	Frequently feel grouchy	.35
7.	Background on social occasions	.28
9.	Burdened by a sense of remorse	.27

The first two items, though they have been classed as indicative of social maladjustment by some *a priori* classifications **(18, 46)**, are classed as symptomatic of depression on physiological bases by Darrow, of emotionality by Bell **(3)**, and their intercorrelations in this study indicate that it is the depressed individual who *says* that he is lonely. The 'shut-in personality' aspect of these symptoms is manifested in the eighth factor. From what has been said above concerning the subjective nature of the stimulus items, it follows that a single item may have different meanings for different individuals. Thus, "lonesome, even with others" may mean depressed to one group of subjects, and shy or timid to another group. The effect of this would be that a single item would function as two distinct items showing high factor loadings on two traits. This is, of course, but one of several possible reasons for an item's having high loadings on more than one trait.

This trait of depression has appeared in previous studies, though

not always formally identified. Bridges (**5**) finds that the items of the Woodworth P. D. Sheet which differentiate delinquent girls from normals are those items which have high loadings on 'D'. Garrison (**15**), using the Thurstone Neurotic Inventory, finds the same result for women prisoners. Garrison's most differentiating items common to this study are 4, 28, 31, 9, and 40. Slawson (**36**) using the Mathews scale reports that the highest consistent differences between his delinquent boys and normals are in items of depression and tendency toward fatigue. Titley (**45**) in a study of the pre-psychotic personality in involutional melancholia, reports asocial trends present. Cameron (**7**) in his study of depression in psychotics says "We have conjectured that the feelings of depression, sadness, *loneliness*, anxiety and apprehension . . . should be regarded as psychic pain due to the action of a stimulus." (Italics added). He also reports that the fluctuation rate of the Necker cube decreases with the depth of depression in psychotics.

The work of Darrow lends confirmation to the interpretation of this trait as depressive in character. All but one of the items listed here as having high loadings on 'D' were classed by Darrow in his 'depressed' category. The physiological measures most consistently associated with these items were low initial resistance, relative impermanence of conditioning and a high proportional change in resistance after shock. A trait of depression was likewise found by Thurstone (**41**) in his analysis of Moore's psychotic symptoms.

The relation of the 'D' factor to the 'C' factor should be noted. Because of the imposition of an orthogonal simple structure, these two traits as here defined are uncorrelated in the population of individuals. However, in the sample of behavior items investigated there is evidence that the same is not true for the universe of behavior items. Items which have high correlations with, and are symptomatic of 'C', tend also to have high loadings with, and to be symptomatic of 'D'. Whether this would obtain in a wider range of behavior items is an hypothesis, but an interesting one, and one which accords with clinical information.

The third trait presents evidence of being identifiable with *hypersensitivity* (using this term metaphorically rather than with any physiological implications) and has been designated 'H'. Its high loadings are:

6.	Feelings easily hurt	.52
19.	Cannot stand criticism	.47

5.	Consider self a nervous person	.45
25.	Often in a state of excitement	.42
15.	Easily discouraged	.40
3.	Worry over humiliating experiences	.38
26.	Frequently feel grouchy	.31
28.	Often just miserable	.30
40.	Cannot make up mind	.26
8.	Ideas run through head at night	.26

The position of item 25 as higher on 'H' than on 'C' is difficult to account for, and no satisfactory explanation has been found. With this single exception, the loadings are consistent with the hypothesis of a trait of 'sensitiveness'. The absence of item 24, ("Touchy on various subjects") proved puzzling until it appeared on a later factor in a position which renders it probable that the item was interpreted to mean: "Touchy on various *college* subjects". This trait is highly similar to Darrow's 'hypersensitive' category, and shows consistent negative relations with recovery reaction quotient, and degree of permanence of conditioning, together with a tendency towards positive relationship with percentage change in resistance after shock. The items which are high on 'H' are among Page's list of schizoid traits, serving to differentiate his normal subjects from manic patients. In the Guilfords' study this trait seems to be contained in their 'E' factor.

The fourth factor may be most parsimoniously termed 'lack of self-confidence'. The general pattern underlying it is *inferiority,* and it has been tentatively designated 'I'. The items which have high loadings on 'I' are:

34.	Lack self-confidence	.66
42.	Not self-confident about abilities	.59
15.	Easily discouraged	.50
39.	Feelings of inferiority	.43
22.	Mind wanders badly	.35
19.	Cannot stand criticism	.31
20.	Difficulty making friends	.30
33.	Self-conscious with superiors	.29
35.	Difficult to speak in public	.28
1.	Stage fright	.25

Individuals possessing high degrees of 'I' are marked by feelings of inferiority, and lack of self-confidence in general, with particular ref-

erence to abilities. This lack of self-confidence is most noticeable in social situations, though the inclusion of more items like 15 and 19 might show that the inferiority was not as confined to social situations as the present results might indicate. No hypothesis can be formed from the data at hand as to the genetic factors underlying this, or, indeed, any of the other factors. That must remain a problem for further investigation. However, Faterson's work (**11**) is suggestive. It is interesting to note that there was no relation between lack of self-confidence and intelligence as measured.

The loadings of the fifth factor are reproduced below:

12.	Troubled with shyness	.71
7.	Keep in background on social occasions	.67
39.	Feelings of inferiority	.60
2.	Difficulty starting conversation	.57
42.	Not self-confident about abilities	.52
34.	Lack self-confidence	.50
21.	Feeling of being watched on street	.48
18.	Bothered by being watched at work	.48
33.	Self-conscious in the presence of superiors	.46
20.	Difficulty making friends	.42
36.	Self-conscious, personal appearance	.34
30.	Hesitate to volunteer in class	.32
40.	Cannot make mind	.31
3.	Worry over humiliating experiences	.30
27.	Self-conscious reciting	.28
10.	Worry over possible misfortune	.25

Two facts are at once striking about this trait. The first is the extent to which it is in agreement with Guilfords' trait 'S', tentatively called by them 'Social Introversion'. This agreement is the more remarkable in that (as has been pointed out) there were but seven items common to the two studies. Such results as these verify the significance of the trait, and serve to validate the method. To quote Thurstone (**42**, p. 55):

"It is a fundamental criterion for a valid method of isolating primary abilities that the weights of the primary abilities for a test must remain invariant when it is moved from one test battery to another test battery."

The results on 'C' and 'S' present an empirical check on the validity of the method by that criterion.

The second fact of significance, which has already been anticipated, is the inter-relationship between 'S' and 'I'. Though measures of these two traits would show that they are uncorrelated in the population of individuals, there is evidence that they are inter-dependent in the universe of behavior items, in that there are many behavior symptoms which are indicative of either trait.

The loadings of items 40 and 10 are higher than might be expected, though they are not so high as to invalidate the hypothesis as to the nature of the trait, nor are they so wholly inconsistent with that hypothesis as to render it untenable. The relatively lower loadings of items 2 and 20 are to be explained by the fact that portions of their variance are already accounted for by other factors. (It is axiomatic that an item of behavior which is symptomatic for two traits cannot be as diagnostic of either as can an item which is a 'pure' measure of one trait alone. Mathematically this is expressed by the condition that the sum of the total variance, including specific and error variance, must be equal to unity.)

The items which are characteristic of this trait are classed together by Darrow in his "social" category. From his published results, none of the physiological measures studied appears characteristic of this trait. Seven of these items appear in Page's list of schizoid traits (**33**), but show no consistent differences between his three groups.

The next trait is interesting in that it appears to differentiate social self-consciousness in a different situation from that described by 'S'. 'S' seems to involve self-consciousness in the more intimate, face-to-face situations where the subject is himself a member of the group, whereas 'P' (which may be taken to stand for 'platform' or 'public-speaking') involves self-consciousness in the more formal situation of speaker-audience relation. This is advanced as an hypothesis subject to verification. This distinction accords with the common experience of finding that the 'good-mixer' may or may not have severe stage fright, and the bashful boy may make a skilled debater. The loadings on 'P' are:

35.	Difficult to speak in public	.80
1.	Stage fright	.77
27.	Self-conscious reciting	.72
30.	Hesitate to volunteer in class	.50

33.	Self-conscious with superiors	.42
12.	Troubled with shyness	.36
2.	Difficulty starting conversation	.35

There is, as would be expected, a certain relation between 'S' and 'P', as shown by the fact that items 2, 12, 33, 27 and 30 have loadings on both factors. The items most diagnostic of one, however, have zero or near-zero loadings on the other. This trait corresponds to none in Guilfords' analysis, though there is reason to suppose that it is included in his 'S' factor and might be separated from it by the inclusion of appropriate items in the battery to be analyzed. 'P' corresponds quite closely to Darrow's category which he designates anxiety. The physiological measure most characteristically associated with the highest items on 'P' is "percentage change in resistance after shock", that measure being correlated positively with four out of the five items highest on 'P'.

The two remaining factors are much less definite than the previous ones have been. They are, however, sufficiently consistent to admit of extremely speculative hypotheses as to their nature, indicating, at least, the direction of further investigation. The seventh factor loadings ('Co' of Table V) are here presented.

23.	Depressed by low marks in school	.55
29.	Bothered by some useless thought	.51
24.	Touchy on various subjects	.35
30.	Hesitate to volunteer in class	.33
38.	Must do a thing over several times	.27
18.	Bothered by being watched at work	.27
15.	Easily discouraged	.25
22.	Mind wanders badly	.25
2.	Difficulty starting conversation	—.24
20.	Difficulty making friends	—.36
43.	Above median, Psychological Examination	—.43

Item 23 may be interpreted as "received low marks in school" rather than as "*depressed* by low marks". That this is the interpretation given the item by the subjects is indicated, not only by the factor loadings, but the actual correlation coefficients themselves. There are three outstanding characteristics of this trait. It may be that the trait is a composite which later analysis will reveal as the linear sum of three traits, intellectual defect, social inadequacy of a particular sort, and

attentive difficulties (items 29, 22, 38, and possibly 43) or that this is a fundamental trait which underlies three seemingly distinct types of behavior. Neither hypothesis can be discarded without further investigation. As to the nature of the trait—or traits—three possibilities occur. The first is that some attentive defect operates to produce low grades, and to produce also a low score on the Psychological Examination, manifesting itself in the useless thought and mind-wandering. A second possibility is that these are samples of behavior symptomatic of an obsessive or compulsive trait. This hypothesis finds difficulty in explaining the negative loading of 'intelligence'. The third, and the one which appears most plausible is that the trait 'Co' represents the personality and behavior items which are caused by a cognitive defect—i.e., by the individual's finding himself below the average of the group in which he is placed. This hpyothesis would consider the low intelligence test score as indicative of basic defect, rather than due to distractibility, leading directly to the low grades in school, the dissatisfaction with school subjects, the feeling of having to do something over, and the annoyance at being watched at work. It explains also the presence of item 30 with the absence of item 27. The hypothesis of cognitive defect experiences difficulty in explaining, however, the loadings on items 29 and 22, and the negative loadings on items 2 and 20 except as compensatory reactions. The question must be submitted to further experimental test of these, and possibly other, hypotheses.

The final factor likewise presents difficulties in interpretation, though the hypothesis of *'austic tendency'* **(34)** is tenable. The 'Au' trait has high loadings on:

13.	Daydream frequently	.62
32.	Periods of loneliness	.49
9.	Burdened by a sense of remorse	.48
36.	Self-conscious, personal appearance	.48
8.	Ideas run through head at night	.47
4.	Lonesome, even with others	.47
21.	Feeling of being watched on street	.46
29.	Bothered by some useless thought	.40
25.	Often in a state of excitement	.34
10.	Worry over possible misfortunes	.32
11.	Happiness to sadness without reason	.27
41.	Ups and downs in mood without cause	.25

The items with highest loadings are consistent with the clinical picture of the shut-in personality; three are represented in Page's list of schizophrenic traits (**33**); two are found diagnostic of schizophrenia by Landis and Katz (**25**) and one other by Smith (**37**). The relationship between 'Au' and 'D' obscures the results of any attempt to determine significant physiological measures from Darrow's study. Before this trait can be accepted as a significant category the hypothesis of 'autistic tendency' will have to be investigated more fully.

Again it must be urged that the behavior items and their loadings define the traits, and such terms as *'cycloid', 'depression', 'hypersensitivity', 'inferiority', 'social introversion', 'public self-consciousness', 'cognitive defect'*, and *'autistic tendency'*, represent merely working hypotheses whose value lies in directing future research into the underlying nature of these concepts which serve to reduce the observed interrelations of a set of forty variables to the simplest set of eight. This investigation is not to be looked upon as the end, but merely as a preliminary approach to the field.

Further research might well take three approaches to the problem. The first, and one which should precede the other two, is to extend the number and the range of behavior situations in which each trait is manifested. Problems would be formulated by selecting an hypothesis as to the nature of a single trait, determining, both *a priori* and empirically under what situations such a trait as that postulated might manifest itself, taking those behavior items as a battery, together with the other known traits as points of reference, and applying the multiple factor techniques. When this has been done and the range of behavior situations in which the trait is manifested is fairly extensive, some attempt at measurement of the trait will be possible. Any attempt to measure the traits as here determined, except a crude approximation, without further research would be premature and unwise. An extension of this method of attack, to be made after the behavior situations are fairly well-defined, is to discover more objectively determinable situations which are symptomatic of the trait. The lead offered by Studman's work (**39**, **40**) for 'C' and by Cameron (**7**) on the fluctuation rate of the Necker cube in 'D' might be highly valuable.

The second attack on the problem, which can better be undertaken after the first is well under way, is the determination of the course of genetic development, and the etiology of each trait. It is, of course, by making possible such investigations that the first approach is valuable. Mere classification and labeling without an honest

attempt to discover further simplifications is sterile. However, before the etiology of a syndrome can be profitably studied, it must first be established that there *is* such a syndrome. This the previous experimental work in the field has failed to do—with, of course, certain all-too-rare exceptions.

A third method of approach which might well supplement the second, but which can scarcely precede the first, is the continuation of Darrow's attempt to discover physiological correlates of these traits—or of the traits that further research might indicate as more fundamental than these.

As to the effect of further research of the first type on these categories, one word might be said. It is quite possible that each of these traits will be discovered to be itself a composite of several others, more basic, just as there is evidence that Guilfords' 'E' has been here broken up into 'C' and 'H', and that 'Co' will likely be found to be composite. It is quite improbable, however, that any two of these traits will be found by further research to be fundamentally one. Such a result could occur only through a highly artificial sampling of behavior items. While the number of dimensions of personality will undoubtedly increase beyond eight, both in this way and by the addition of entirely new concepts, it can be safely asserted that it is at least eight, even when personality is restricted to the 'single trait' of neurotic tendency.

Summary and conclusions—

In an attempt to investigate the concept of neurotic tendency, 39 of the 42 most discriminative items in the Thurstone Neurotic Inventory were administered as a questionnaire to a group of 500 male college students. The responses of 'Yes' or 'No' were considered as forcing a continuous variable into a dichotomy and the tetrachoric correlation coefficients between each item and every other item, and between each item and intelligence were computed graphically. The test-retest consistency coefficients for each item were computed, the retest occurring after an interval of one week. The table of intercorrelations was factored by the centroid method of Thurstone, and eight factors found sufficient to account for the observed intercorrelations with negligible residuals. The eight centroid factors were then transformed into a simple structure. From the results the following conclusions seem justified:

1. There is no single trait of neurotic tendency which can be postulated in a parsimonious description of behavior. The current questionnaires which purport to measure the single factor of neurotic tendency are composites of at least eight traits.

2. Tentative hypotheses have been formed as to the nature of these primary traits. Cycloid tendency, depression, hypersensitiveness, lack of self-confidence, social maladjustment and self-consciousness in the speaker-audience situation appear to be well-verified as hypotheses. Autistic tendency and cognitive defect are offered as possible explanations for the two remaining traits.

3. The consistency of response to the individual items on two occasions a week apart is high. When this is used as an approximation to the reliability coefficient, the proportion of the variance of each item which remains unaccounted for (specific variance) has a median value of .23.

4. The results of this experiment show consistencies with other empirical attempts to determine behavior categories.

TABLE I

QUESTIONNAIRE ITEMS

1. Do you get stage fright?
2. Do you have difficulty in starting conversation with a stranger?
3. Do you worry too long over humiliating experiences?
4. Do you often feel lonesome, even when you are with other people?
5. Do you consider yourself a rather nervous person?
6. Are your feelings easily hurt?
7. Do you keep in the background on social occasions?
8. Do ideas run through your head so that you cannot sleep?
9. Are you frequently burdened by a sense of remorse?
10. Do you worry over possible misfortune?
11. Do your feelings alternate between happiness and sadness without apparent reason?
12. Are you troubled with shyness?
13. Do you day-dream frequently?
14. Have you ever had spells of dizziness?
15. Do you get discouraged easily?
16. Do your interests change quickly?
17. Are you easily moved to tears?
18. Does it bother you to have people watch you at work, even when you do it well?
19. Can you stand criticism without feeling hurt?
20. Do you have difficulty making friends?
21. Are you troubled with the idea that people are watching you on the street?
22. Does your mind often wander badly so that you lose track of what you are doing?
23. Have you ever been depressed because of low marks in school?
24. Are you touchy on various subjects?
25. Are you often in a state of excitement?
26. Do you frequently feel grouchy?
27. Do you feel self-conscious when you recite in class?
28. Do you often feel just miserable?
29. Does some particular useless thought keep coming into your mind to bother you?
30. Do you hesitate to volunteer in a class recitation?
31. Are you frequently in low spirits?
32. Do you often experience periods of loneliness?

33. Do you often feel self-conscious in the presence of superiors?
34. Do you lack self-confidence?
35. Do you find it difficult to speak in public?
36. Do you often feel self-conscious because of your personal appearance?
37. If you see an accident are you quick to take an active part in giving help?
38. Do you feel that you must do a thing over several times before you leave it?
39. Are you troubled with feelings of inferiority?
40. Do you often find that you cannot make up your mind until the time for action has passed?
41. Do you have ups and downs in mood without apparent cause?
42. Are you in general self-confident about your abilities?
43. (Above the median, A. C. E. Psychological Examination.)

TABLE II
PERCENTAGE FREQUENCY OF NEUROTIC RESPONSE TO EACH ITEM

Item Number	Description	Percentage Frequency
1.	Stage fright	.610
2.	Difficulty starting conversation	.286
3.	Worry over humiliating experiences	.458
4.	Lonesome, even with other people	.388
5.	Consider self a nervous person	.270
6.	Feelings easily hurt	.360
7.	Keep in background on social occasions	.336
8.	Ideas run through head at night	.346
9.	Burdened by a sense of remorse	.246
10.	Worry over possible misfortune	.354
11.	Happiness to sadness without reason	.374
12.	Troubled with shyness	.270
13.	Day-dream frequently	.658
15.	Easily discouraged	.300
16.	Interests change quickly	.450
18.	Bothered by being watched at work	.442
19.	Cannot stand criticism	.206
20.	Difficulty making friends	.102
21.	Feel watched on street	.158
22.	Mind wanders badly	.322
23.	Depressed by low marks in school	.668
24.	Touchy on various subjects	.394
25.	Often in a state of excitement	.248
26.	Frequently feel grouchy	.356
27.	Self-conscious reciting	.424
28.	Often feel just miserable	.228
29.	Bothered by some useless thought	.290
30.	Hesitate to volunteer in class	.476
31.	Frequently in low spirits	.240
32.	Periods of loneliness	.332
33.	Self-conscious with superiors	.456
34.	Lack self-confidence	.330
35.	Difficult to speak in public	.570
36.	Self-conscious, personal appearance	.338
38.	Must do a thing over	.360
39.	Troubled with feelings of inferiority	.290
40.	Cannot make up mind	.384
41.	Ups and downs in mood	.416
42.	Not in general self-confident	.264
43.	Above median of freshman class Psychol. Exam.	.538

TABLE III

CENTROID FACTOR LOADINGS

Item Number	Centroid Loadings I	II	III	IV	V	VI	VII	VIII
1	448*	—513	328	140	212	—100	292	080
2	365	—550	—142	—132	097	106	143	—134
3	453	159	009	—025	—081	251	—062	—225
4	507	230	—424	—085	264	—191	—187	285
5	410	170	036	—125	215	135	139	—256
6	438	259	159	038	143	245	037	—102
7	383	—442	—272	—312	—119	153	095	051
8	391	221	—089	—250	156	—037	—191	—168
9	457	426	—130	—053	116	—113	—076	—097
10	456	259	—125	—129	—165	—081	095	—163
11	452	468	—175	283	—191	—244	333	—125
12	501	—631	—020	—099	—120	066	032	—090
13	356	206	040	045	111	—286	—409	—111
15	557	091	215	225	—053	274	—177	143
16	293	264	175	198	—152	081	—121	—099
18	405	—215	232	—272	—125	149	044	—063
19	329	166	303	110	155	267	—181	136
20	378	—414	—446	099	179	310	131	224
21	638	—132	—140	—098	—050	—125	—160	—184
22	302	192	134	244	—200	—107	—152	128
23	285	117	335	—255	—208	—132	145	178
24	426	154	161	—079	—050	—031	—033	177
25	371	316	177	117	261	—072	046	—247
26	512	297	—114	189	177	058	094	048
27	443	—494	315	—030	112	—237	187	108
28	562	457	—195	—164	082	120	211	132
29	394	369	205	—192	—257	—202	—126	028
30	356	—386	327	—229	009	—093	088	195
31	627	353	—257	017	054	068	250	209
32	577	357	—360	—231	223	—160	—102	257
33	536	—434	127	061	—019	—105	—059	—155
34	626	—417	—119	258	—115	181	—130	239
35	399	—560	264	171	219	—251	251	119
36	455	—161	—053	—119	015	—231	—211	—201
38	215	073	069	—161	—210	123	082	072
39	640	—322	—067	071	—241	050	—183	—048
40	558	073	120	002	—132	078	043	—086
41	493	524	—281	283	—241	—108	296	—129
42	418	—479	—126	184	—269	113	—250	183
43	139	—035	—235	230	210	137	—182	—289

*Decimal points properly preceding each entry have been omitted.

TABLE IV
MATRIX OF TRANSFORMATION TO SIMPLE STRUCTURE

Cen-troid Factor	Primary Trait C	D	H	S	I	P	Co	Au
I	392*	342	326	285	332	488	171	404
II	436	172	261	—473	—240	—530	267	278
III	—111	—503	355	413	057	—215	615	—093
IV	402	—253	000	045	682	—373	—381	—156
V	—379	355	472	383	—087	—399	—416	146
VI	—078	065	626	—422	094	275	—040	—577
VII	550	150	—044	442	—467	—009	—085	—505
VIII	—172	619	—288	044	360	—246	441	—343

*Decimal points properly preceding each entry have been omitted.

TABLE V
PRIMARY TRAIT MATRIX

Item Number	Loadings on Primary Traits C	D	H	I	S	P	Co	Au
1	048*	026	131	252	234	766	017	—098
2	—079	118	075	048	572	350	—235	—045
3	250	020	380	067	302	—117	064	183
4	073	679	006	181	026	—032	—032	467
5	206	129	452	—156	163	121	—034	199
6	251	115	518	077	065	032	112	119
7	—057	282	—068	001	668	128	—043	—067
8	026	200	260	—089	155	—055	042	467
9	298	271	236	—011	—010	—076	050	481
10	409	151	083	—103	248	—046	113	316
11	808	141	—011	010	—057	013	—012	274
12	—045	009	003	218	708	362	—052	002
13	023	018	106	221	—054	000	075	620
15	202	098	397	498	154	012	254	075
16	289	—125	240	221	000	—106	179	168
18	001	—013	206	003	475	206	274	020
19	010	097	471	307	—057	016	244	046
20	002	500	067	299	420	126	—361	—228
21	142	131	080	170	479	139	—013	461
22	262	—011	—001	349	—053	—031	248	190
23	165	066	008	—090	104	203	549	066
24	158	200	143	142	094	102	346	176
25	290	001	417	—004	—134	174	—006	342
26	387	346	310	187	—026	052	—036	199
27	—026	033	—009	180	278	720	149	037
28	431	560	298	—092	099	—069	158	190
29	252	033	047	—002	063	—057	507	403
30	—140	084	020	096	319	503	328	—018
31	513	591	200	085	116	016	081	142
32	161	714	088	035	061	—045	100	490
33	045	—092	080	293	459	419	—018	223
34	100	236	046	655	499	172	—018	—030
35	005	040	—022	282	184	799	—036	—036
36	—003	022	023	109	344	184	—019	475
38	147	081	067	—032	214	—041	267	—038
39	145	042	041	425	595	113	045	209
40	317	053	256	144	306	106	196	180
41	849	199	038	037	014	—128	—035	246
42	—033	066	—134	585	516	049	004	—018
43	017	—019	228	174	083	—101	—428	172

*Decimal points properly preceding each entry have been omitted.

TABLE VI
ANALYSIS OF RELIABILITY, COMMUNALITY, AND SPECIFICITY

Item Number	r_{jj} Reliability	h_j^2 Communality	s_j^2 Specificity
1	.95	.738	.212
2	.91	.532	.378
3	.87	.355	.515
4	.88	.719	.161
5	.97	.363	.607
6	.93	.378	.552
7	.92	.563	.357
8	.88	.363	.517
9	.58	.451	.129
10	.74	.377	.363
11	.87	.757	.113
12	.88	.687	.192
13	.91	.447	.463
15	.85	.545	.305
16	.87	.279	.591
18	.89	.382	.508
19	.83	.386	.444
20	.93	.718	.212
21	.91	.531	.379
22	.82	.296	.524
23	.90	.386	.514
24	.85	.273	.577
25	.86	.419	.441
26	.79	.445	.345
27	.88	.656	.224
28	.83	.673	.157
29	.77	.494	.276
30	.92	.490	.430
31	.79	.698	.092
32	.84	.795	.045
33	.82	.534	.286
34	.91	.766	.144
35	.93	.760	.170
36	.86	.388	.472
38	.81	.153	.657
39	.88	.619	.261
40	.90	.364	.536
41	.86	.851	.009
42	.78	.635	.145
43	—	.308	—

Omit decimals

REFERENCES

1. ALLPORT, G. W. *Review.* Character & Pers., 1934, **3**, pp. 168-169.
2. ALLPORT, G. W., and ODBERT, H. S. *Trait Names.* Psychol. Monogr., 1936, **47**, No. 211, pp. 171.
3. BELL, H. M. *Manual for the Adjustment Inventory.* Stanford University: Stanford Univ. Press, 1934.
4. BERNREUTER, R. G. "Theory and Construction of the Personality Inventory." *J. Soc. Psychol.*, 1933, **4**, pp. 387-405.
5. BRIDGES, J. W. "A Study of a Group of Delinquent Girls." *Ped. Sem. & J. Genet. Psychol.*, 1927, **34**, pp. 187-204.

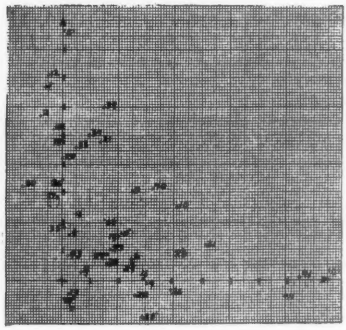


FIGURE 1

6. BURKS, B. S. "Personality Theories in Relation to Measurement." *J. Soc. Psychol.*, 1936, **7**, pp. 140-150.
7. CAMERON, D. E. "Studies in Depression." *J. Ment. Sci.*, 1936, **82**, pp. 148-161.
8. CHANT, S. N. F. "Multiple Factor Analyses and Psychological Concepts." *J. Educ. Psychol.*, 1935, **26**, pp. 263-272.
9. CHESIRE, L., SAFFIR, M., and THURSTONE, L. L. *Computing Diagrams for the Tetrachoric Correlation Coefficient.* Chicago, Univ. of Chicago Bookstore, 1933, p. 57.
10. DARROW, C. W. *Reaction Tendencies Relating to Personality.* in Stone, C. P., Darrow, C. W., Landis, C., & Heath, L. L. *Studies in the Dynamics of Behavior.* Chicago: Univ. of Chicago Press, 1932, pp. xiv, 332.
11. FATERSON, H. A. "Organic Inferiority and the Inferiority Attitude." *J. Soc. Psychol.*, 1931, **2**, pp. 87-101.
12. FLANAGAN, J. C. "Technical Aspects of Multi-trait Tests." *J. Educ. Psychol.*, 1935, **26**, pp. 641-651.
13. ———— *Factor Analysis in the Study of Personality.* Stanford University: Stanford Univ. Press, 1935, pp. x, 103.
14. FRANZ, S. I. *Handbook of Mental Examination Methods.* New York: Macmillan, 1919, p. 193.
15. GARRISON, K. C. "Study of Emotional Characteristics of Women Prisoners." *J. Soc. Psychol.*, 1932, **3**, pp. 472-477.
16. GUILFORD, J. P. "Examination of a Typical Test of Introversion-Extroversion by Means of the Method of Similar Reactions. *J. Soc. Psychol.*, 1933, **4**, pp. 430-443.
17. GUILFORD, J. P., and GUILFORD, R. B. "Personality Factors *S, E,* and *M* and Their Measurement." *J. Psychol.*, 1936, **2**, pp. 109-127.
18. HARVEY, O. L. "Concerning the Thurstone Personality Schedule." *J. Soc. Psychol.*, 1932, **3**, pp. 240-251.
19. HEIDBREDER, E. "Measuring Introversion and Extroversion." *J. Abnorm. Soc. Psychol.*, 1926, **21**, pp. 120-134.
20. HUMM, D. G., and WADSWORTH, G. W. "The Humm-Wadsworth Temperament Scale." *Amer. J. Psychiat.*, 1935, **92**, pp. 163-200.
21. JOHNSON, W. B. "The Effect of Mood on Personality Traits As Measured by Bernreuter." *J. Soc. Psychol.*, 1934, **5**, pp. 515-522.
22. KELLEY, T. L. *Statistical Method.* New York: Macmillan, 1923, pp. 253-258.
23. LAIRD, D. A. "Detecting Emotional Behavior." *J. Abnorm. Soc. Psychol.*, 1925, **20**, pp. 128-141.
24. LANDIS, C. "Questionnaires and Study of Personality." *J. Nerv. Ment. Dis.*, 1935, **81**, p. 345. (Abstract)

25. Landis, C., and Katz, S. E. "Validity of Questions Which Purport to Measure Neurotic Tendency." *J. Appl. Psychol.*, 1934, **18**, pp. 343-356.
26. Line, W., Griffin, J. D., and Anderson, G. W. "The Objective Measurement of Mental Stability." *J. Ment. Sci.*, 1935, **81**, pp. 61-106.
27. Lotka, A. J. *Elements of Physical Biology.* Baltimore: Williams & Wilkins, 1925, pp. xxx, 460.
28. Maller, J. B. Character Sketches. *A Test for the Measurement of Personality Adjustment.* New York: Columbia University, 1932.
29. Moore, T. V. Series of research publications of Ass'n for Research in Nerv. & Ment. Dis., 1930, No. 9, 324-339.
30. Moore, T. V. "The Empirical Determination of Certain Syndromes Underlying Praecox and Manic-Depressive Psychoses." *Amer. J. Psychiat.*, 1930, **9**, pp. 719-738.
31. Mosier, C. I. "A Note on the Criteria of Internal Consistency." *Psychometrika*, 1936, **1**, No. 4.
32. Mosier, C. I. *A Factor Analysis of Certain Neurotic Symptoms.* Ph.D. thesis, Univ. of Chicago Library, 1937.
33. Page, J., Landis, C., and Katz, S. E. "Schizophrenic Traits." *Amer. J. Psychiat*, 1934, **13**, pp. 1213-1225.
34. Rosanoff, A. J. *Manual of Psychiatry.* New York: 1927, Wiley, pp. 122-124, 335-336.
35. Russell, O. R. "Some Observations on Multiple-Factor Analysis." *J. Educ. Psychol.*, 1935, **26**, 284-290.
36. Slawson, J. "Psychoneurotic Responses of Delinquent Boys." *J. Abnorm. Soc. Psychol.*, 1925, **20**, pp. 261-281.
37. Smith, H. N. "Neurotic Tendencies Shown in Dementia Praecox and Manic-Depressive Insanity." *J. Soc. Psychol.*, 1933, **4**, pp. 116-128.
38. Smith, R. B. "Measurement of Inferiority Feelings at the High School Level." *Psychol. Bull.*, 1932, **29**, 633. (Abstract).
39. Studman, L. G. "Studies in Experimental Psychiatry. V. '*W*' and '*f*' Factors in Relation to Traits of Personality." *J. Ment. Sci.*, 1935, **81**, pp. 107-137.
40. Studman, L. G. "Factor Theory in the Field of Personality." *Character & Pers.*, 1935, **4**, pp. 34-43.
41. Thurstone, L. L. "The Vectors of Mind." *Psychol. Rev.*, 1934, **41**, pp. 1-32.
42. Thurstone, L. L. *Vectors of Mind.* Chicago: Univ. of Chicago Press, 1935, pp. xv, 266.
43. Thurstone, L. L. "The Bounding Hyperplanes of a Configuration of Traits." *Psychometrika*, **1**, 1936, pp. 61-68.
44. Thurstone, L. L. and Thurstone, T. G. "A Neurotic Inventory." *J. Soc. Psychol.*, 1930, **1**, pp. 3-30.
45. Titley, W. B. "Pre-Psychotic Personality in Involutional Melancholia." *J. Nerv. Ment. Dis.*, 1935, **82**, p. 559.
46. Willoughby, R. R. "Some Properties of the Thurstone Personality Schedule and a Suggested Revision." *J. Soc. Psychol.*, 1932, **3**, pp. 401-424.
47. Willoughby, R. R. "Norms for the Clark-Thurstone Inventory." *J. Soc. Psychol.*, 1934, **5**, pp. 91-97.
48. Wrightstone, J. W. "Validity of the Woodworth-Mathews Personal Data Sheet." *J. Educ. Psychol.*, 1934, **25**, pp. 39-44.

PSYCHOMETRIKA—VOL. 2, NO. 4
DECEMBER, 1937

BUROS, OSCAR K. *Educational, Psychological and Personality Tests of 1936*, Rutgers University Bulletin, Vol. 14, No. 24. Studies in Education, No. 11. New Brunswick, N. J.: School of Education, Rutgers University, August, 1937. P. 141. $0.75. Paper.

A REVIEW

This manual, a supplement to *Educational, Psychological and Personality Tests of 1933, 1934 and 1935*, includes 353 tests (and other devices) published in 1936 and those omitted from the earlier publication, excepting certain classes which have been excluded from both issues. As the compiler recognizes, non-paper-and-pencil tests are inadequately represented. For each test is given the following information: title, appropriate age or grade level, date, individual or group, number of forms, cost, time, author, publisher, and references (when known). Judgments as to the relative quality of the tests included are not given.

A new section lists 291 books in the field of measurement, accompanied by book-review excerpts from various sources. Some of the excerpts might have been improved had the compiler paraphrased more freely instead of extracting passages verbatim. However, the ready accessibility of the opinions of several reviewers is a distinct advantage. Following the book section are indexes of authors of the tests and books, publishers of measurement books in English-speaking countries, periodicals containing the reviews which are cited, and titles of the tests and books listed.

This manual, having been compiled with accuracy and thoroughness, is, on the whole, admirably adapted to its purpose. To test technicians and to those school administrators and psychologists who are interested in the testing field, Buros is offering a genuine service.

DOROTHY C. ADKINS,
The University of Chicago.

MONROE, WALTER G. and ENGELHART, MAX D. *The Scientific Study of Educational Problems.* New York; The Macmillan Company, pp xv + 504.

A REVIEW

This book, well adapted to the improvement of research on educational problems, represents an interesting mixture of, and compromise between, the imparting of practical wisdom and incisive treatment of abstract theory. The practical admonitions on elementary matters to the tyro educationist may well produce a chuckle; the mirth cannot exceed a chuckle, because anyone acquainted with the literature of the field knows these simple directions are all too necessary. The authors clearly know what is necessary. However, many of the more advanced statistical techniques are presented, with careful attention to their application

to educational research. The absence of derivations for many of these formulas is, in part, compensated for by critical evaluation of their applicability to educational and psychometric problems. The authors are fully aware of the present status of educational research, and their book offers much that is distinctly in advance of the field in general.

M. W. RICHARDSON.

KARL J. HOLZINGER assisted by FRANCES SWINEFORD and HARRY HARMAN. *Student Manual of Factor Analysis.* Prepared at the Statistical Laboratory, Department of Education, The University of Chicago, 1937. vi + 101 pp.

A REVIEW

The first two chapters deal with rather general aspects of factor analysis while the next three give a detailed account of the author's bi-factor method, with a hypothetical example completely worked out. The treatment in these five chapters is so detailed and simple that it can be followed by anyone with a thorough knowledge of elementary statistics.

The sixth and seventh chapters deal briefly with other methods of multiple-factor analysis and the relationship between these methods and the bi-factor method. The theory of boundary conditions is briefly presented in the concluding chapter. These three chapters assume a background of higher algebra, matrix theory, and solid analytical geometry.

HAROLD GULLIKSEN.

DECEMBER, 1937
PSYCHOMETRIKA—VOL. 2, NO. 4

INDEX FOR VOLUME 2

AUTHORS

AUTHORS

AUTHORS

TOPICS

In the selection of the articles to be printed in Psychometrika, an effort will be made to attain objectivity of choice. All papers will be received by one person who will first remove from the article the name of contributor and institution. This person will then send it to three or more persons who will make independent judgments upon the suitability of the article submitted. This procedure seems to offer a possible mechanism for making judicious and fair selections. In order to facilitate decisions on articles, and to safeguard anonymity, the following suggestions are made to contributors:

1. If possible, submit three copies of the article. Rough sketches of completed drawings will suffice for the two extra copies.
2. Do not include name and professional connection on the article, but write these on separate sheet containing title of article, and seal in an envelope. Paste this envelope on the last page of the article. Avoid references which might reveal the identity of the author. Place such references, if necessary, in the envelope.
3. Send articles to

M. W. RICHARDSON,
Managing Editor,
Psychometric Corporation
5822 Drexel Avenue
Chicago, Illinois.

Application for membership should be sent to
IRVING D. LORGE
Chairman of the Membership Committee
Institute of Educational Research
Teachers College
Columbia University
New York City.

Psychometrika is issued four times per year. Quarterly issues appear in March, June, September, and December. The subscription price to libraries and other institutions is $10.00 per year. This price includes extra clean copies for binding. The annual dues of $5.00 from each member include subscription to the journal. Such dues are payable to the Treasurer. Subscriptions should be sent to the editorial and business office.

Editorial and Business office:
Psychometric Corporation,
5822 Drexel Avenue
Chicago, Illinois

The officers of the Society for the year 1937-1938 are as follows:

President: J. P. GUILFORD, University of Nebraska, Lincoln, Nebraska.

Secretary: JACK W. DUNLAP, Catherine Strong Hall, University of Rochester, Rochester, New York.

Treasurer: ALBERT K. KURTZ, 173 Cornwall Street, Hartford, Connecticut.

PRINTED BY THE DENTAN PRINTING CO., COLORADO SPRINGS, COLO.